Charles University
Faculty of Science

Study program: Biology

Branch of study: Immunology

Petr Tauš

Experimental verification of *in silico* predicted protein binder to FOXO4 transcription factor and transcriptome analysis of bladder cancer

Experimentální ověření *in silico* predikovaného vazebného proteinu k transkripčnímu faktoru FOXO4 a analýza transkriptomu nádorů močového měchýře

Diploma thesis / Diplomová práce

Supervisor: RNDr. Karel Drbal, Ph.D.

Prague, 2017

Prohlášení:

Prohlašuji, že jsem tuto diplomovou práci zpracoval samostatně na základě konzultací se školitelem a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 14. 8. 2017 Petr Tauš

Poděkování:

Na tomto místě chci v prvé řadě poděkovat mému školiteli Karlu Drbalovi za odborné konzultace, podporu a trpělivost. Mé díky dále patří Petru Novákovi a Evženu Bouřovi za umožnění provedení velké části experimentů v jejich laboratořích. Dále bych rád poděkoval Jiřímu Vondráškovi za provedení *in silico* predikce, Lukáši Slavatovi a Danu Kavanovi za cenné rady v oblasti exprese a purifikace rekombinantních proteinů a Palle Villesen za poskytnutou příležitost vypracovat bioinformatickou část této diplomové práce v jeho laboratoři. V neposlední řadě bych rád poděkoval všem svým skvělým kolegům a kamarádům, kteří mi byli vždy ochotní poskytnout cenné rady a drobnou pomoc. Děkuji!

Abstrakt:

Tato diplomová práce je rozdělená na experimentální a bioinformatickou část, které jsou spolu provázány přes transkripční faktory z rodiny ‘forkhead box O‘ (FOXO). FOXO mají klíčovou roli v mnoha buněčných procesech, jako například v regulaci buněčného cyklu, apoptózy a metabolismu. Dlouhou dobu byly považovány pouze za nádorové supresory, ale vzrůstající počet studií poukazuje i na jejich pronádorovou roli. Z tohoto důvodu jsou intenzivně studovány jako potenciální terapeutické cíle v nádorových onemocněních. *In silico* predikce protein-protein interakcí se v poslední dekádě těší vzrůstající oblibě jak v základním výzkumu, tak při vývoji nových léčiv. Nicméně výsledky v mnoha případech stále nejsou dostatečně přesné při srovnání s očekávanými vlastnostmi predikovaných biomolekul. V této práci jsem pomocí termoforézy v malém měřítku ověřil vazbu čtyř *in silico* predikovaných vazebných proteinů, založených na přirozeně se vyskytující PDZ doméně k transkripčnímu faktoru FOXO4.

Neinvazivní nádory močového měchýře představují heterogenní onemocnění, u kterého se i přes veškeré snahy stále nedaří přesně predikovat agresivita daného nádoru. V bioinformatické části této práce jsem na základě transkriptomických dat z pacientských vzorků nádorů močového měchýře popsal složení nádorového mikroprostředí a určil prognostickou hodnotu jednotlivých buněčných typů v předpovědi progrese onemocnění. Dále jsem popsal vyšší úroveň parazánětu vycházejícího z nádorových buněk, který překvapivě asocioval s neprogresivními nádory. U těchto nádorů jsem dále zjistil vyšší expresi FOXO1 a naopak nižší expresi FOXO6.

Klíčová slova:

alternativní vazebné proteiny, nádory močového měchýře, předpověď progrese nádorů, FOXO, FOXO4, NMIBC, parazánět, PDZ, RNA-Seq, nádorové mikroprostředí, xCell

Abstract:

This diploma thesis includes an experimental and a bioinformatic part. The two parts are linked together through the subject of transcription factors of 'forkhead box O' (FOXO) family. FOXO transcription factors have a key role in many cellular processes including cell cycle regulation, apoptosis and metabolism. For a long time, they have been considered strictly as the tumor-suppressors yet a growing number of evidence is pointing out to their pro-tumorigenic role. In consequence FOXO transcription factors are studied intensively as potential therapeutic targets in cancer. In the past decade, in silico prediction of protein-protein interactions has become popular in basic research as well as in drug development. Nonetheless, the predicted structures are still far from fitting to the expected behavior of the respective biomolecules. In the experimental part of this thesis, I verified the interaction of four *in silico* predicted protein binders based on naturally occurring PDZ domain with FOXO4 using microscale thermophoresis.

Non-invasive bladder tumors represent a heterogeneous disease where reliable prediction of tumor aggressiveness is still lacking despite an intensive research. In the bioinformatic part of this thesis, I described the cellular composition of the tumor microenvironment and demonstrated its prognostic value based on transcriptomic data from patient samples of bladder tumors. Surprisingly, I showed higher level of para-inflammation, which stems from cancer cells, associated with non-progressive tumors. Additionally, higher expression of FOXO1 and lower expression of FOXO6 was present in these tumors.

Keywords:

alternative binding proteins, bladder cancer, cancer progression prediction, FOXO, FOXO4, NMIBC, para-inflammation, PDZ, RNA-Seq, tumor microenvironment, xCell

Table of contents

Glossary:

BCa	Bladder cancer
DBD	DNA-binding domain
FOX	Forkhead box
GSEA	Gene set enrichment analysis
IEC	Ion exchange chromatography
IMAC	Immobilized metal ion affinity chromatography
MDS	Multidimensional scaling
MIBC	Muscle invasive bladder cancer
MST	Microscale thermophoresis
nanoDSF	Nano differential scanning fluorimetry
NMIBC	Non-muscle invasive bladder cancer
NN	Neural network
PFS	Progression free survival
PI	Para-inflammation
PPI	Protein-protein interactions
PTMs	Post-translational modifications
RF	Random forest
SEC	Size exclusion chromatography
SVM	Support vector machine
TME	Tumor microenvironment
Trp	Tryptophan

Abbreviations of cell types:

CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
GMP	Granulocyte-macrophage progenitor
HSC	Hematopoietic stem cells
MEP	Megakaryocyte-erythroid progenitor
MPP	Multipotent progenitor
MSC	Mesenchymal stem cells

1 Introduction

BCa is the 5th most common cancer in Central Europe with a high recurrence rate and with approximately one in four patients developing advanced cancer. Due to this fact, patients are monitored frequently and often undergo multiple resection surgeries over many years. BCa is a very heterogeneous disease and current therapeutic approaches do not benefit all patients. Our inability to accurately predict which case of BCa will progress to an advanced invasive stage and to stratify patients for particular treatments results in one of the highest economic and emotional burdens among cancers. Therefore, development of accurate prediction models for patient monitoring and surveillance is urgently needed.

The 'O' group of the large forkhead box (FOX) transcription factor family consists of four members (FOXO1/3/4/6). The whole FOX family shares a conserved 'winged helix'/'forkhead' DNA binding domain. FOXO proteins play an important role in the regulation of key cellular processes such as the cell cycle, apoptosis and metabolism. Their main physiological role is to maintain homeostasis under various stress conditions. Therefore, the dysregulation of FOXO function often results in many pathologies such as cancer. In the past years, they were considered only as tumor suppressors but nowadays a large number of evidence supports their context-dependent role in cancer. The suggested explanation for their paradoxical tumor-promoting role is that they may regulate homeostasis of both normal and cancer cells. In this thesis, I analyzed their expression levels in RNA-Seq data obtained from patient samples after bladder cancer (BCa) surgery with respect to progression and differences in the tumor microenvironment (TME).

A growing body of evidence supports the key role of TME and especially tumor-infiltrating lymphocytes in tumorigenesis. Several attempts have been made to leverage immunological biomarkers as a tool for the prediction of prognosis and for stratifying patient treatment. Amongst other prognostic tools, the most promising appeared to be 'Immunoscore' that is based on the numeration of two lymphocyte populations and their location within the tumor. Traditional experimental approaches describing tumor-infiltrating leukocytes within TME are cytometry and immunohistochemistry, however, they are limited with the amount of available markers. In this thesis, I used the emerging *in silico* method called xCell for dissecting TME from transcriptomic data. To create prediction models for disease progression and to characterize cell type prognostic value I employed the Random Forest (RF), Support Vector Machine (SVM) and Neural Network (NN) machine learning algorithms. Additionally, I identified low-grade inflammation, termed as

para-inflammation (PI), in tumors based on a 40-gene signature. Cancer cells themselves are the origin of PI and it is associated with adverse outcomes in most cancer types but not in BCa.

Specific protein-protein interactions (PPI) are the major structural and regulatory mechanisms for the functioning of all living organisms. Their study at the molecular level is crucial for the elucidation of cellular and immune functions, knowledge of which is necessary for the development of new therapeutic methods. The investigation of these interactions is experimentally difficult and often impossible because of the extreme complexity of interactions in the eukaryotic system, especially in mammalian cells. One of the extreme reductionist approaches for studying PPI is structural *in silico* modelling.

During the last decade, *in silico* prediction of PPI became a widely used method in primary research as well as in drug design. However, the predicted structures are still far from fitting to their *in vivo* behavior of the respective biomolecules. Experimental data are essential for the improvement of the prediction algorithms.

In our project we chose the mouse PDZ domain (PDB ID: 3VQF) as a suitable structure for rational design (Figure 8). We identified the key interface residues in the complex of the PDZ domain and a biologically relevant target – transcription factor FOXO4 (PDB ID: 1E17) and we selected ten mutations in the PDZ domain resulting in expected higher interaction affinity between the two molecules. To determine the protein stability of the predicted PDZ mutant variants and their binding affinity to FOXO4 I used nano differential scanning fluorimetry (nanoDSF) and microscale thermophoresis (MST) respectively.

In this thesis, I proved that the PDZ domain structure is suitable for rational design and verified our choice of these mutant variants, however, for the reasons that I explained in the section ‘The aims of the project’ I could not have finished the planned project. Therefore, this thesis additionally contains a bioinformatic part that is interconnected with the experimental part via the FOXO transcription factors. In the bioinformatic part, I illuminated the prognostic landscape of the cellular composition of TME and PI in BCa using bioinformatic tools. My findings will help in further experimental investigation and will become a foundation for single cell analysis of patient samples by flow cytometry in my home laboratory.

1.1 Bladder cancer

BCa is the 5th most common cancer in Central and Western Europe and the 9th most common cancer worldwide. In 2015, there were ~541 000 newly diagnosed patients with bladder cancer and ~188 000 deaths. BCa is more prevalent in men (1 in 59 men vs 1 in 239 women being diagnosed before age 79 years) and in countries with a high socio-demographic index (Fitzmaurice et al., 2017).

Most BCa are urothelial carcinomas, which start in the urothelial cells that line the inside of the bladder. On the basis of their morphology they can be divided into papillary, solid and mixed types. The most common type is the papillary type, particularly in non-muscle invasive BCa (NMIBC). NMIBC are in ~50 % cases low-grade, while most muscle-invasive BCa (MIBC) are high grade. Histologic grade is the most important prognostic factor in NMIBC. It is based on pathologist assessment how cancer cells look under the microscope. If they look like normal cells then the tumor is denoted as low-grade meaning well-differentiated tumor and vice versa. In BCa diagnostic, two different grading systems are used by pathologists. The first, introduced in 1973 by WHO (1973 WHO) is a numerical grading system based on cellular anaplasia (Mostofi et al., 1973). The second, 2004 WHO is a categorical grading system based on architectural and cytological atypia (Figure 1) (Eble et al., 2004). The extent of tumor invasion is described by Tumor-Node-Metastasis system, which is the most important prognostic factor in MIBC (Figure 1). The majority of BCa at time of diagnosis (~60 %) are Ta stage non-muscle invasive tumors, usually of low

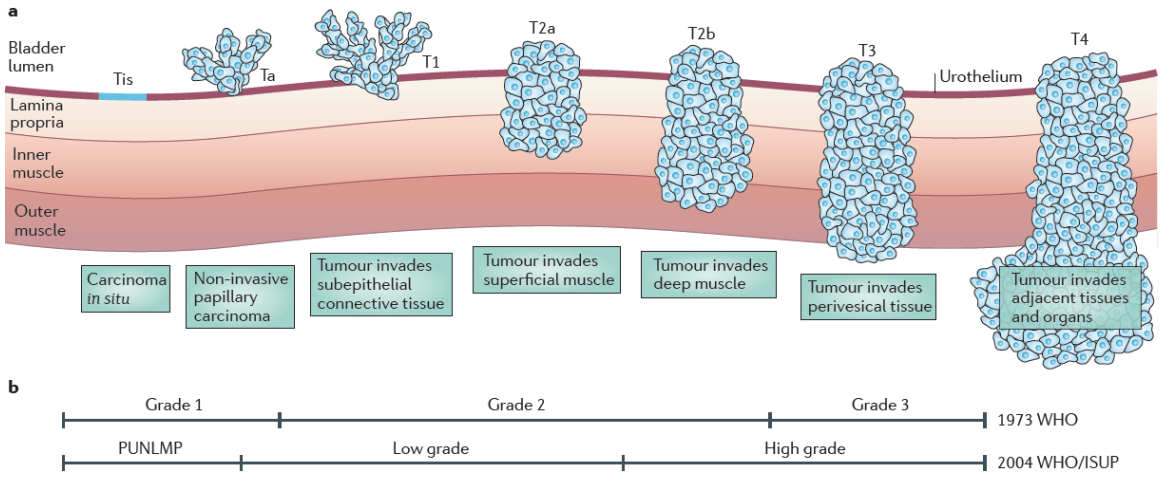


Figure 1 | **Staging and grading of bladder cancer.** A | BCa stages according to the Tumor-Node-Metastasis system. NMIBC are up to stage T1, MIBC are T2 stage and higher. B | BCa 1973 WHO and 2004 WHO/International Society of Urological Pathology grading system. Adapted from (Knowles and Hurst, 2015).

grade. Approximately 20 % of tumors at diagnosis are T1 stage that already penetrated the epithelial basement membrane but have not invaded muscle tissue. T1 stage tumors are generally of high grade. The rest of the tumors at diagnosis (~20 %) are MIBC that frequently metastasize with a five-year survival rate < 50 %. Ta and T1 NMIBC have a high rate of recurrence (50-70 %), a relatively low rate of progression to MIBC (10-15 %) and ~90 % five-year survival rate (Knowles and Hurst, 2015).

The heterogeneity of NMIBC results in widely different outcomes. Patients with NMIBC are monitored with cystoscopies and due to a high rate of recurrence they often have to undergo multiple resection surgeries over many years. Consequently the economic and emotional burden of BCa is higher compared to other cancer types (Hong and Loughlin, 2008). Even though risk tables with prognostic tools are available (Kluth et al., 2015) their prediction accuracy is rather low and therefore accurate multimarker models for stratifying patient treatment and application of therapeutic drugs are urgently needed.

1.1.1 Molecular classes of bladder cancer

It was shown that BCa can be grouped into papillary or non-papillary on the basis of different genetic alterations. These are activating mutations in FGFR3 in papillary tumors and inactivating mutations in tumor suppressor genes TP53 and RB1 in non-papillary tumors. Both subtypes show a high frequency of mutations in genes encoding chromatin-modifying enzymes, with characteristic differences between them (Eich et al., 2017).

Choi et al. (2014) defined three distinct molecular genomic subtypes of MIBC. The first termed basal-like are characterized by p63 activation, positive CK5/6, EGFR and CD44 expression, and lack of CK20. In addition, this subtype is more aggressive than others. Next, luminal subtypes possess an active estrogen receptor pathway, ERB2 and PPARγ expression profiles, and are typically enriched in activating FGFR3 mutations. Third, p53-like subtypes are characterized by the wild-type TP53 gene expression signature. These three subtypes differ in their responsiveness to systemic chemotherapy. Basal-like appeared to be sensitive to neoadjuvant chemotherapy, whereas p53-like was associated with resistance to neoadjuvant MVAC therapy. Interestingly, all chemoresistant tumors displayed a p53-like phenotype after therapy.

Recently, Hedegaard et al. (2016) demonstrated that NMIBC can be subgrouped into three distinct molecular classes with different clinical outcomes (Figure 2). These

classes possess basal- and luminal-like characteristics similar to those of MIBC (Choi et al., 2014). Class 1 contained mainly Ta tumors of low grade with luminal-like characteristics. Tumors within class 2 were luminal-like too but this class represented a high-risk subclass with the most progression events among molecular classes. Tumors within class 1 were characteristic for high expression of early cell-cycle genes. On the contrary, tumors within the class 2 were characteristic for high expression of late cell-cycle genes. Furthermore, increased activity of epithelial-mesenchymal transition and stem-like features were associated with the class 2. Tumors within class 3 demonstrated a basal-like gene expression pattern that has not been previously described in NMIBC. They clearly differed from tumors within class 1 and 2 with low cell-cycle and metabolic activity. The authors suggested that the class 3 represents tumors in a dormant state (Hedegaard et al., 2016).

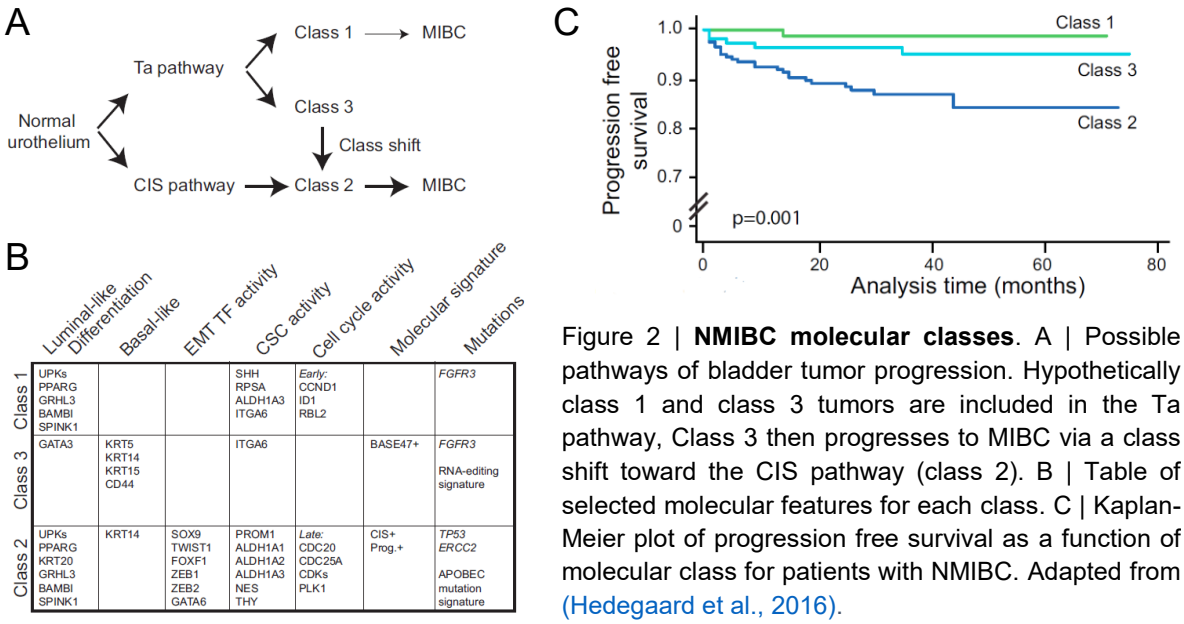


Figure 2 | **NMIBC molecular classes**. A | Possible pathways of bladder tumor progression. Hypothetically class 1 and class 3 tumors are included in the Ta pathway, Class 3 then progresses to MIBC via a class shift toward the CIS pathway (class 2). B | Table of selected molecular features for each class. C | Kaplan-Meier plot of progression free survival as a function of molecular class for patients with NMIBC. Adapted from (Hedegaard et al., 2016).

Several analyses of MIBC revealed distinct genomic subtypes and clusters (reviewed in (Eich et al., 2017)). However, these molecular subtypes only partly overlap. Thus, a meeting at the Spanish National Cancer Research Center was organized in 2016 to seek a consensus on a BCa molecular taxonomy. A strong consensus was reached regarding the existence and terminology of tumors coined as basal-squamous-like tumors. These are typical by high expression of KRT5/6 and KRT14, and low to undetectable expression of FOXA1 and GATA3. They also agreed on the existence of a second molecular subtype of MIBC enriched for FGFR3 gene alteration and GATA3, FOXA1 and

KRT20 expression. Nevertheless, more work has to be done for more reliable evidence of the remaining molecular subtypes (Eich et al., 2017).

1.2 Introduction to FOXO transcription factors

The FOX protein family of transcription factors shares an evolutionarily conserved DNA binding 'winged helix'/'forkhead' domain consisting of approximately 100 amino acids (Figure 3). To date, more than 40 FOX proteins have been discovered in the human genome ((http://www.genenames.org/cgi-bin/genefamilies/ set/508 (accessed 7.22.17)), for the FOX gene nomenclature see (Kaestner et al., 2000)). They are grouped in 19 subclasses in alphabetical order from A to S class. In this thesis, I focused only on the 'O' subclass (the human orthologs of Drosophila dFOXO) that contains four members (FOXO1, FOXO3, FOXO4 and FOXO6) that primarily differ in their tissue-specific expression. The best-studied member of this family is FOXO1 and the least one is FOXO6. FOXO proteins also differ in their size. FOXO1 (655 aa) and FOXO3 (673 aa) are bigger than FOXO4 (505 aa) and FOXO6 (492 aa).

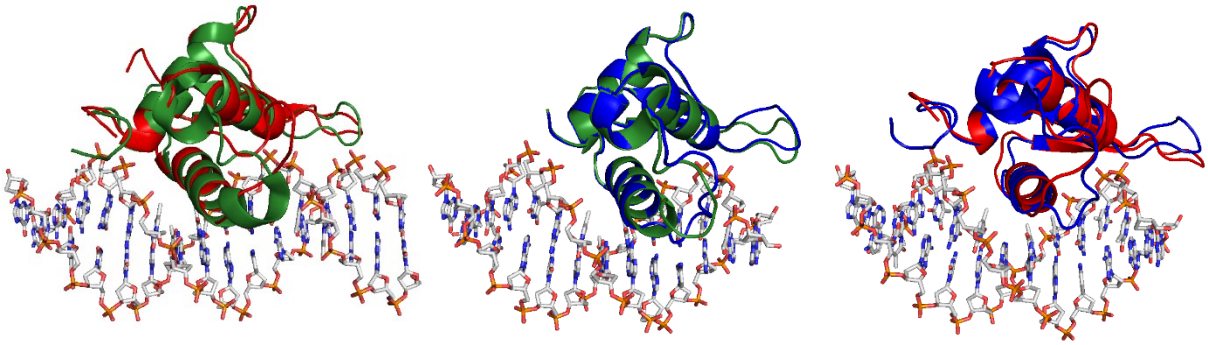


Figure 3 | **Structural alignment of FOXO-DNA-binding domain (DBD)** | Structural alignment of FOXO1-DBD (green), FOXO3-DBD (red) and FOXO4-DBD (blue) in complex with dsDNA. Alignment and visualization were created in PyMOL software based on PDB ID 3CO6 (FOXO1), 2UZK (FOXO3) and 3L2C (FOXO4).

1.2.1 FOXO4 structure

In the experimental part of this thesis I primarily focused on FOXO4. Its DNA-binding domain (DBD) was chosen for targeting by PDZ domain and its mutant variants that were predicted *in silico*. FOXO4-DBD consists of approximately 110 amino acids and is composed of three alpha-helices (H1, H2 and H3), three beta-sheets (S1, S2 and S3) and two wing-like loops (W1 and W2). Their topological arrangement in DBD is H1-S1-H2-H3-S2-W1-S3-W2. Specific DNA contacts are mediated mainly by H3 helix, which is anchored in the large DNA groove. The N-terminus and W1 wing are in contact with

phosphate groups of DNA. The loop between H2 and H3 helices is also involved in the binding between FOXO4-DBD and DNA but not in other FOXO proteins (Figure 4). Additionally, based on the structure of the complex of DBD-FOXO4 and its dsDNA consensus sequence (13 bp), it was suggested that uniquely arranged water molecules around DBD-FOXO4-dsDNA complex contribute to the binding specificity (Boura et al., 2010).

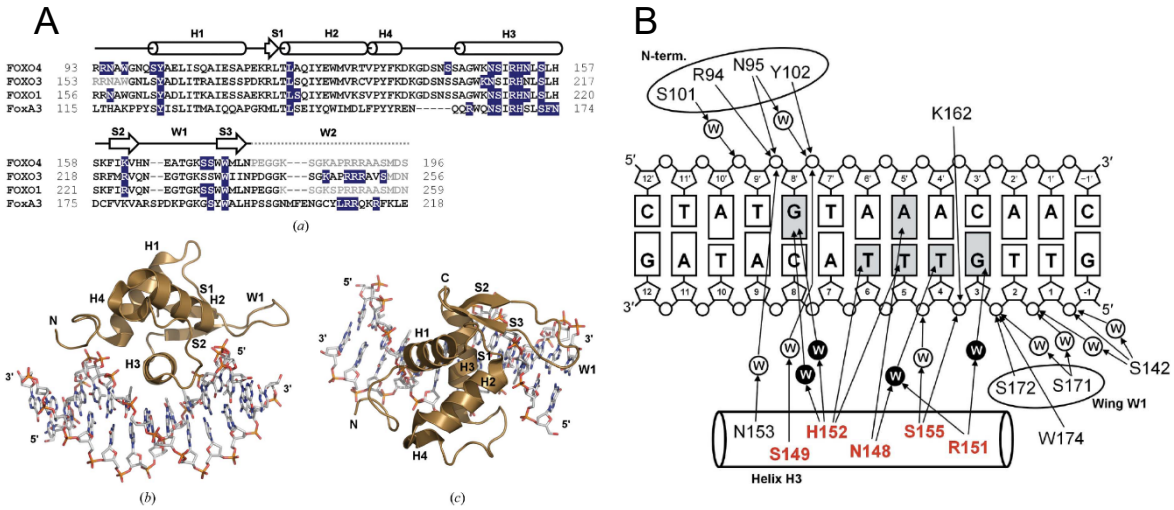


Figure 4 | **FOXO4.** A | (a) Sequential alignment of DBD of FOXO 4/3/1 and FoxA3. Arrangement of the secondary structure showed above. Blue-labeled amino acids mediate protein-DNA interaction. Gray labelled missing in the corresponding structural complex. (b) Structure of the FOXO4-DNA complex with labelled elements of the secondary structure. (c) The rotated structure of fig (b) 90° towards the reader around the x-axis. B | Schematic representation of the contacts between DNA and FOXO4-DBD. Polar interactions are highlighted in red and gray. Water molecules (W) in black circles participate in specific interactions between bases and FOXO4-DBD. Adapted from (Boura et al., 2010).

1.2.2 FOXO regulation

FOXO activity is regulated at multiple levels mainly by transcriptional regulations, post-transcriptional modifications and post-translational modifications (PTMs). The examples of transcription factors which directly regulate the expression of FOXO genes are p53, E2F-1 and FOXO themselves (reviewed in (Urbánek and Klotz, 2016; Klotz et al., 2015)). PTMs such as phosphorylation, acetylation and ubiquitination affect FOXO subcellular localization, DNA binding, FOXO stability and interaction with transcriptional coregulators. FOXO are substrates for many kinases (recent list of kinases can be found in (Klotz et al., 2015)) .

For example, the serine-threonine protein kinase B (AKT) is the key regulator of FOXO activity (Figure 5) (Brunet et al., 1999). All FOXO contain three AKT phosphorylation sites except for FOXO6 that contains only two (Jacobs et al., 2003).

Activation of the PI3K/AKT pathway by insulin or insulin-like growth factor 1 results in the interaction of FOXO proteins with 14-3-3 proteins and subsequent cytoplasmic sequestration of FOXO-14-3-3 complexes (reviewed in (Tzivion et al., 2011)). This applies for FOXO1/3/4 but not for FOXO6 which is mainly nuclear and its activity is not controlled by subcellular localization (Jacobs et al., 2003).

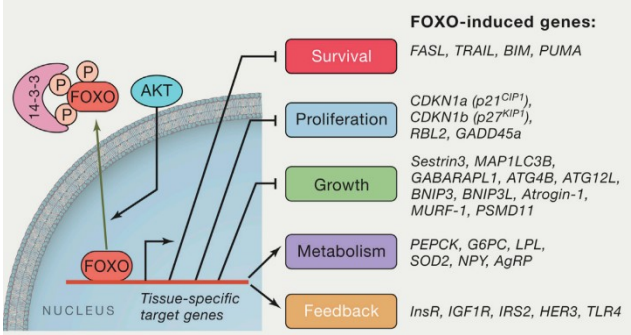


Figure 5 | **AKT-mediated FOXO regulation.** AKT phosphorylation of FOXO proteins generates recognition motifs for the 14-3-3 family of phosphor-binding proteins, which consequently leads to sequestration of phosphorylated FOXO in the cytosol. It results in the attenuation of the expression of its target genes, a partial list of which is shown. Adapted from (Manning and Toker, 2017).

Other AGC family kinases (Pearce et al., 2010) can phosphorylate AKT phosphorylation sites too. For instance, it was shown that serum/glucocorticoid inducible kinase (SGK; another important downstream component of PI3K signaling pathway) phosphorylates FOXO3 at the AKT phosphorylation sites (Brunet et al., 2001). Phosphorylation by both AKT and SGK inhibits FOXO function. However, it was demonstrated that phosphorylation can also lead to activation of FOXO proteins. For example, Jun N-terminal kinase (JNK) and protein kinase R-like ER kinase (PERK) phosphorylate FOXO4 (Essers et al., 2004) and FOXO1 (Zhang et al., 2013) respectively and promote their function by stimulating their nuclear accumulation. Similarly to phosphorylation, acetylation and ubiquitination of FOXO proteins can both promote or inhibit FOXO transcriptional activity too (reviewed in (Klotz et al., 2015; Daitoku et al., 2011)). A complex nature of PTMs prompted Calnan and Brunet (2008) to suggest the theory of the 'FOXO code' where enzymes would be writers of the PTMs and FOXO binding proteins the readers. The 'FOXO code' would then represent the overall combinations of PTMs that would determine the type of elicited cellular responses. However, at that time little was known about post-transcriptional regulations and now an accumulating body of evidence shows their important role in regulation of FOXO proteins (reviewed in (Urbánek and Klotz, 2016)). Post-transcriptional regulations are more complex and context dependent than regulations by PTMs. It makes this topic even harder to solve but at the same time it opens new doors towards possible therapeutic interventions.

1.2.3 FOXO pathways

FOXO proteins recognize and bind to the 5‘-GTAAA(T/C)AA-3‘ (or 5’-TT(A/G)TTTAC-3’) consensus sequence and with lower affinity to the 5‘-(C/A)(A/C)AAA(C/T)AA-3‘ consensus sequence known as DAF-16 family-binding element (Furuyama et al., 2000) and insulin-responsive element (Guo et al., 1999) respectively (reviewed in (Obsil and Obsilova, 2011)). They are involved in the regulation of cellular processes such as the cell cycle, apoptosis, metabolism and oxidative stress resistance (reviewed in (Eijkelenboom and Burgering, 2013)). Their biological role is mainly to maintain homeostasis and respond to stress conditions. A demonstrated role of FOXO proteins in stem cell maintenance highlighted their homeostatic function over time (Tothova et al., 2007).

In addition, FOXO transcription factors possess several cell type-specific downstream effects in the immune system (reviewed in (Dejean et al., 2011)). For example, FOXO play important roles throughout T cell biology (reviewed in (Hedrick et al., 2012)) and in innate immune homeostasis (Becker et al., 2010). FOXO proteins are also important key regulators of inflammation, for instance, Foxo1 was described to increase the expression of proinflammatory Tlr4 signaling pathway genes in mature macrophages

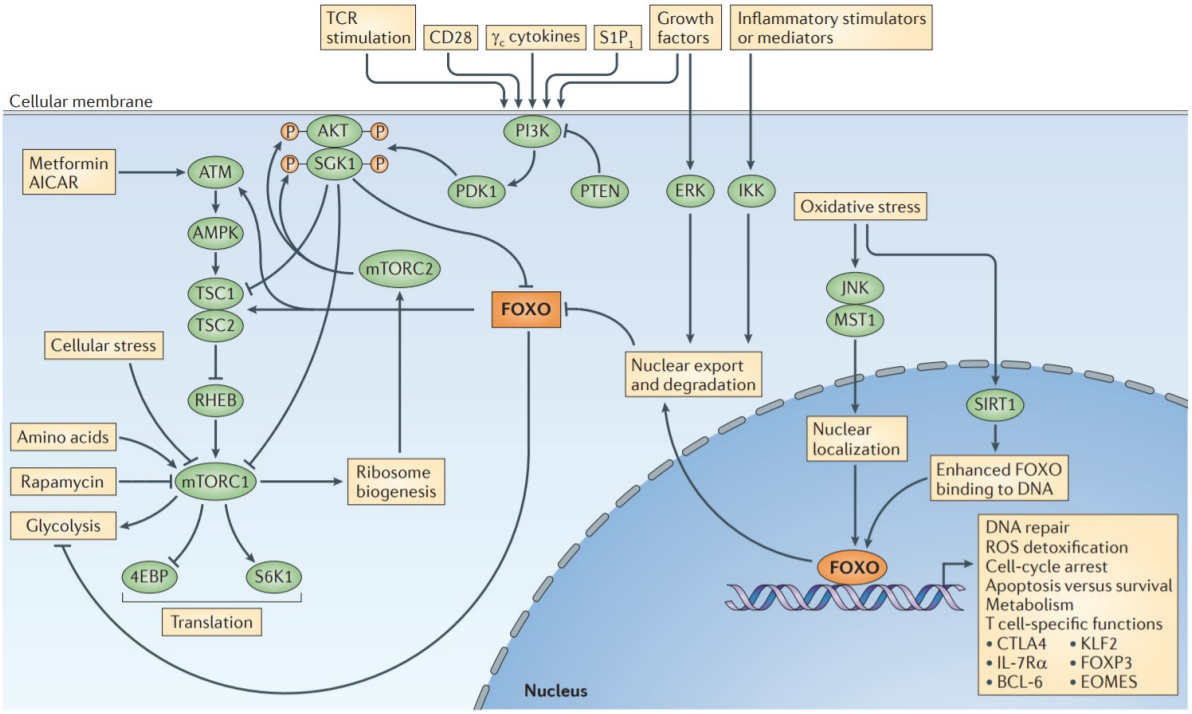


Figure 6 | **A signalling scheme of FOXO transcription factors regulation.** FOXO interact with multiple cellular signalling networks. Examples of such interactions include: PI3K and AKT, JNK, IKK, mTOR and ERK. Adapted from (Hedrick et al., 2012).

(Fan et al., 2010) . Recently Foxo1-mediated autophagy was shown to be vital for NK cell development (Wang et al., 2016) and the role of Foxo1 in regulating alveolar macrophages polarization toward type 2 immune response in asthma was demonstrated (Chung et al., 2016).

FOXP3 transcription factor belonging to the FOX protein family is known to be a master regulator in the development of regulatory T cells (Treg) for a decade (Hori et al., 2003). It was also proven that another member of FOX protein family Foxo1 has an important role in controlling Treg development and function. Its deletion in a murine model led to lymphoproliferative diseases. Additionally, the absence of Foxo3 worsens the effect of the deletion of Foxo1 (Kerdiles et al., 2010). In a recent study Luo et al. (2016) emphasized the importance of Foxo1 in Treg. They showed that Foxo1 plays a crucial role in inhibiting the activated phenotype of Treg and importantly they demonstrated that tumor infiltrating Treg are more sensitive to Foxo1-triggered depletion than Treg from normal tissue. In addition to Foxo1 and Foxo3, Fu et al. (2017) recently described an important role of Foxo4 in regulation of Treg anti-inflammatory function in arthritis. They showed that Foxo4 and Stat3 promote IL-10 expression in response to the anti-inflammatory growth factor progranulin.

A growing body of evidence demonstrates the important roles of FOXO transcription factors (especially FOXO1) in the regulation of the immune system. I assume that a deeper knowledge of their function in a context-dependent manner will help to unveil the relationship between inflammation, metabolism and oxidative stress and how these conditions influence an immune response.

1.2.4 FOXO in cancer

FOXO transcription factors integrate diverse signaling pathways in response to environmental stimuli and translate them into gene expression programs that influence a variety of physiological functions. Therefore it is not surprising that their dysregulation can lead to many pathological processes such as cancer and insulin resistance-related metabolic disorders. The latter mentioned was associated with a higher risk of different types of cancer and there is emerging evidence that dysregulation of FOXO proteins could provide a link between these two pathologies (reviewed in (Link and Fernandez-Marcos, 2017)).

The role of FOXO proteins in cancer and their potential therapeutic targeting have been intensively studied for the last decade (Brachène and Demoulin, 2016; Lam et al.,

2013; Myatt and Lam, 2007). They possess anticancer activity by activating genes involved in apoptosis and cell cycle arrest and for a long time they were considered only as tumor suppressors. For example, inactivation or downregulation of FOXO4 protein was linked to progression of gastric cancer (L. Su et al., 2014), prostate cancer (B. Su et al., 2014), colon cancer (Kwon et al., 2010), cholangiocarcinoma (Lee et al., 2009) and breast cancer (S. H. Kim et al., 2009). However, a growing amount of literature showed that the role of FOXO proteins in cancer development and therapy is context dependent and they may act as a double-edged sword. The plausible explanation of this paradoxical outcome is their role in regulation of homeostasis of both normal and cancer cells. Increased FOXO activity was observed in approximately 40 % of patients with acute myeloid leukemia regardless of the genetic subtype. Inactivation of Foxo1/3/4 in a murine model of MLL-AF9-induced myeloid leukemia led to lower disease burden, disrupted function of leukemia-initiating cells and improved animal survival (Sykes et al., 2011). Huo et al. (2014) demonstrated that FOXO1/3/4 promote hepatoma cell proliferation through transcriptional activation of Insulin-like growth factor 1 receptor (IGF1R). It was also shown that FOXO 1/3/4 promote survival of hematopoietic stem cells with DNA damage caused by chemotherapeutic agents such as cisplatin (Lei and Quelle, 2009). Recently Pan et al. (2017) described how PI3K or AKT inhibitors or taxanes (a class of widely used anticancer agents) induced FOXO1 nuclear localization that led to chemotherapy resistance in cancer cells. Additionally, a recent study showed that FOXO can support tumor survival in a more intricate way. A key metabolic enzyme isocitrate deyhodrogenase 1 (IDH1) is often mutated in cancer cells and when mutated it produces D-2-hydroxuglutarate, a putative oncometabolite that stimulates cancer progression. FOXO1, FOXO3 and with lower effect FOXO4 were demonstrated to redundantly promote expression of IDH1 and thereby maintain the level of D-2-hydroxuglutarate in tumors (Charitou et al., 2015).

In recent years, FOXO proteins have emerged as potential targets for inhibition in certain types of cancer, particularly in chemotherapy-resistant tumors. Their further exploration and consequent elucidation of the role they play in tumorigenesis will be of medical importance.

1.2.5 FOXO in bladder cancer

The altered expression of several members of the FOX protein family has been associated with clinical outcomes and with distinct molecular classes of BCa. For a recent review see (Yamashita et al., 2017). Here I focused only on members of the FOXO group and their implication in bladder cancer.

Shiota et al. (2010) associated low expression of FOXO3 with increased BCa invasiveness and poor overall survival. They investigated these associations and found out that FOXO3 promoted the expression of E-cadherin ('cell adhesion molecule') and negatively regulated two genes involved in cancer cell invasiveness (TWIST1 and Y-box-binding protein). Additionally, they demonstrated that FOXO knockdown led to increased mobility of BCa cell lines. On the basis of the study results the authors suggested that FOXO3 could be a negative regulator of epithelial-mesenchymal transition in BCa (Shiota et al., 2010). Interestingly, increased activity of epithelial-mesenchymal transition, which is characterized by high expression of FOXF1 (besides other genes) was observed in molecular subclass 2 (Figure 2).

Lloreta et al. (2017) analyzed the relationship between FOXO1 expression and the clinical outcome of BCa and the link between the alterations of p53 and FOXO1 expression. Their findings correspond to the previous study which associated FOXO1 expression with grade, stage, recurrence, progression and survival for the first time (T.-H. Kim et al., 2009). Additionally, they showed a strong association between combination of FOXO1 downregulation/p53 overexpression and high grade/stage and demonstrated that only the more aggressive tumors carried both changes (Lloreta et al., 2017).

1.3 Tumor microenvironment

Besides cancer cells, tumors are composed of diverse cell types that can additionally differ in activation states. This widely heterogeneous cellular composition of tumors, coined as the tumor microenvironment (TME), has been intensively studied in recent years by various techniques. Major attention of researchers was attracted by the tumor-infiltrating lymphocytes which are thought to play the most important role in cancer immunity. However, they are only a part of an incredibly complex and dynamic tumor milieu which consists of multiple immune, stromal and other cell types that interact with each other and most importantly with cancer cells. TME is nowadays recognized as a key player in tumor initiation, progression and metastasis as well as in tumor growth inhibition and is in the spotlight of current cancer research (Aran et al., 2017; Gentles et al., 2015; Hanahan and Coussens, 2012).

For example, tumor infiltrating $CD8^+$ T cells and $CD45RO^+$ memory T cells possess antitumor functions and have been associated with a positive prognostic value in many cancer types. This knowledge is now being translated into clinical practice as 'Immunoscore' (Fridman et al., 2012; Galon et al., 2006). However, when considering

other immune cells, such as regulatory T cells or $CD4^+$ T cells, their prognostic effects often differ between cancer types (Gentles et al., 2015; Fridman et al., 2012). In a recent study Gentles et al. (2015) exploited publicly available bulk tumor gene expression data to analyze the relationship between survival and differences in the amounts of 22 distinct tumor-infiltrating leukocytes. $\gamma\delta$T cells and in general an increased level of T cells were found to associate with increased survival rates while myeloid cells and polymorphonuclear (PMN) cells were associated with lower survival rates in solid non-brain tumors. Their results also showed that prognostic values of analyzed cell types are context dependent which corresponds to the previous findings.

In this thesis, I implemented the tools for digital dissection of the TME using bulk tumor gene expression data. In general there are two types of these techniques – one is based on gene set enrichment associated with the individual cell types (e.g. xCell (Aran et al., 2017)) and another is based on algorithmic deconvolution method (e.g. CIBERSORT (Newman et al., 2015)). These analytical algorithms are primarily being developed to overcome limitations of traditional approaches for analysis of cellular heterogeneity (e.g. flow cytometry and immunohistochemistry) such as the limited number of available markers and the need for dissociation of solid tissues which can distort cell representation. Secondarily they are aimed to leverage a huge amount of both RNA-Seq and microarray expression data that are publicly available.

1.4 Para-inflammation in cancer

Inflammation is now recognized as a dominant feature and a well-established hallmark of cancer (Hanahan and Weinberg, 2011). The term PI (*para*- the Greek prefix for near) was coined by R. Medzhitov (2008). He described it as an adaptive response to persistent tissue stress that has characteristics that are intermediate between tissue homeostasis and an inflammatory state. Its physiological function is to restore homeostasis and can be induced by DNA damage or mutations. If tissue stress lasts for a sustained period, PI can become chronic with adverse effects. Aran et al. (2016) created a 40-gene signature (see gene signature in Supplementary information) for analyzing PI from bulk RNA-Seq expression data. They demonstrated that the source of PI are cancer cells and not the immune cells within TME. This distinguishes PI from other cancer-promoting inflammatory responses. PI score significantly differed between cancer types but even within some cancer types. Pan-cancer analysis of 18 cancer types from The Cancer Genome Atlas showed a negative correlation between PI score and survival rate

with the rare exception of chromophobe renal cell carcinoma, prostate cancer and bladder cancer – all tumors of urinary tract (Aran et al., 2016).

1.5 PDZ domain

The PDZ structure typically contains an antiparallel beta-barrel composed of 5-6 beta-sheets and 1-2 alpha-helices. In most cases, PDZ domains recognize a PDZ-binding motif located on elongated C-ends of proteins. Less often, they can also recognize internal binding motifs, lipids or other PDZ domains (Luck et al., 2012). Approximately 18 years ago, Schneider et al. (1999) made various variants of the PDZ domains by mutagenesis and subsequent *in vivo* selection, which specifically targeted new peptides. A recently published study described the creation of a bivalent so-called PDZbody targeting human papillomavirus E6 oncoprotein (with nanomolar affinity), which naturally interacts with PDZ-containing proteins. PDZ domain affinity towards E6 oncoprotein was increased by phage display and adding an alpha-helix from another PDZ domain. Then the PDZbody was formed by joining these PDZ domains (Karlsson et al., 2015).

The PDZ domain displays the properties of a newly occurring class of high affinity protein binders, which are often referred to in literature as alternative binding proteins (Figure 7). These are developed on the basis of small (typically from 5 to 20 kDa) rigid protein structures with high thermal and hydrodynamic stability, on which a new binding site is created. This is the opposite of today's commonly used antibodies based on immunoglobulin fold. They are referred to as alternative because their major goal is to overcome limitations of the antibodies and become their full-fledged alternative in basic research as well as in industry and clinical use. The drawbacks of antibodies stem mainly from their structure and biophysical properties. Their size around 155 kDa limits tissue penetration and prevents targeting of hidden epitopes. Although significant progress has been made in the field of rational design in recent years (Dunbar et al., 2016; Lapidoth et al., 2015), the prediction of new antibody binding sites remains an intricate and unresolved issue, primarily because of the complicated and flexible structure of antibody CDR binding loops. Furthermore, limited and expensive production in eukaryotic expression systems, time and money consuming mutagenic and selection systems, high immunogenicity and limitations of the intellectual property are the disadvantages that can be overcome by the development of the alternative binding proteins (Owens, 2017; Skrlec et al., 2015; Banta et al., 2013).

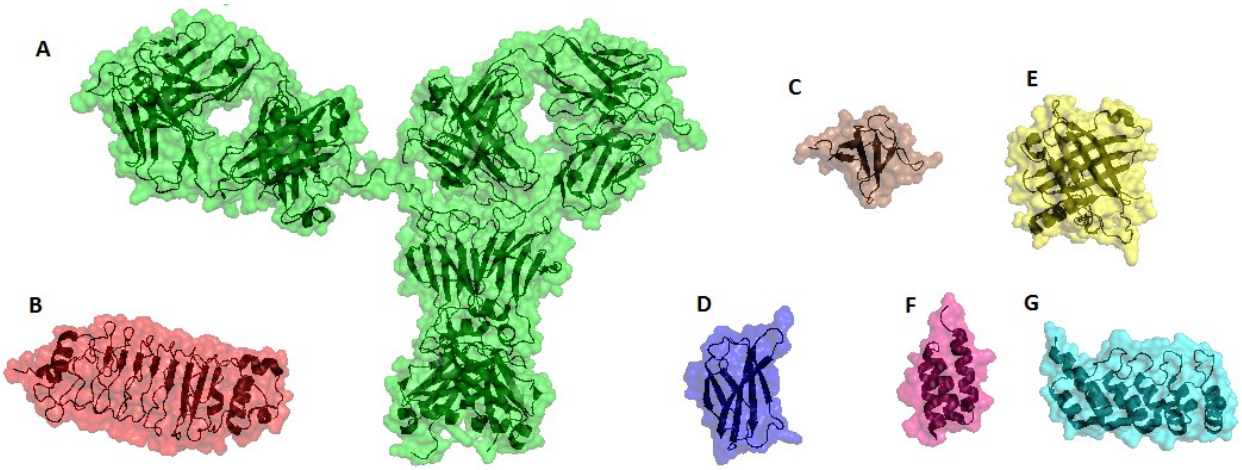


Figure 7 | **Size comparison of IgG and alternative binding protein structures.** A | IgG ~ 155 kDa; B | Repebody ~ 29 kDa; C | Fynomer ~ 7 kDa; D | Adnectin ~ 10 kDa; E | Anticalin ~ 20 kDa; F | Affibody ~ 6 kDa; G | DARPins ~ 17 kDa. Visualization was created in PyMOL software based on PDB ID 1HZH, 4J4L, 1M27, 4JE4, 4GH7, 3MZW and 4HRL respectively.

The selected PDZ domain (PDB ID: 3VQF) naturally found in E1 mouse ubiquitin-protein ligase Lex1 is a 84 amino acid long protein (ca 10 kDa) and its structure contains two alpha-helices and five beta-sheets (Figure 8). The predicted binding site for DBD-FOXO4 predominantly consists of two rigid beta-sheets and flexible loop (Figure 10). Ten mutations were designed within this area to increase the PDZ binding affinity for DBD-FOXO4 (Figure 9). In general, these are mutations optimizing electrostatic binding but there are also two mutations to hydrophobic amino acids. To the best of my knowledge, this presented binding site on the PDZ domain has not yet been described in literature.

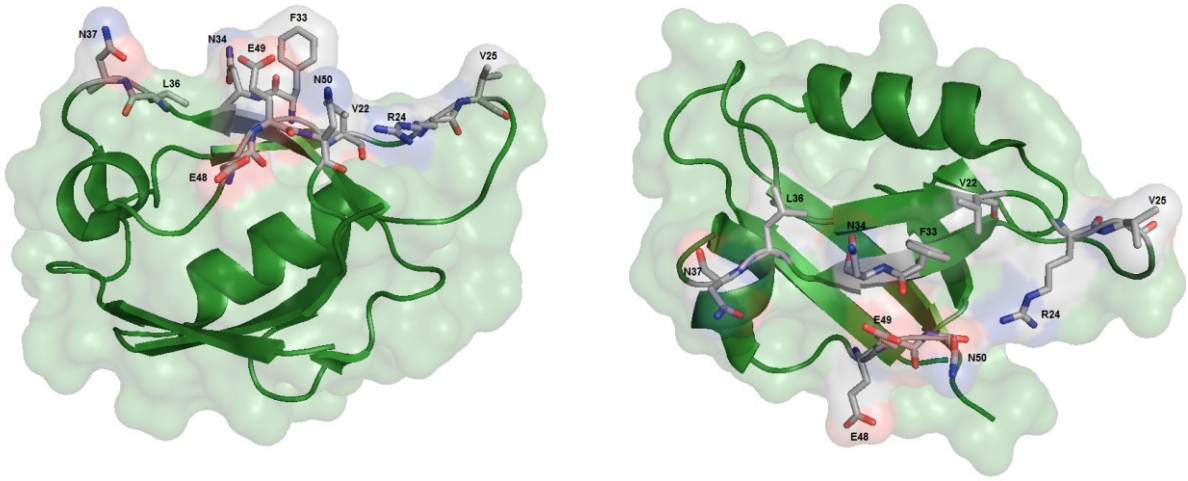
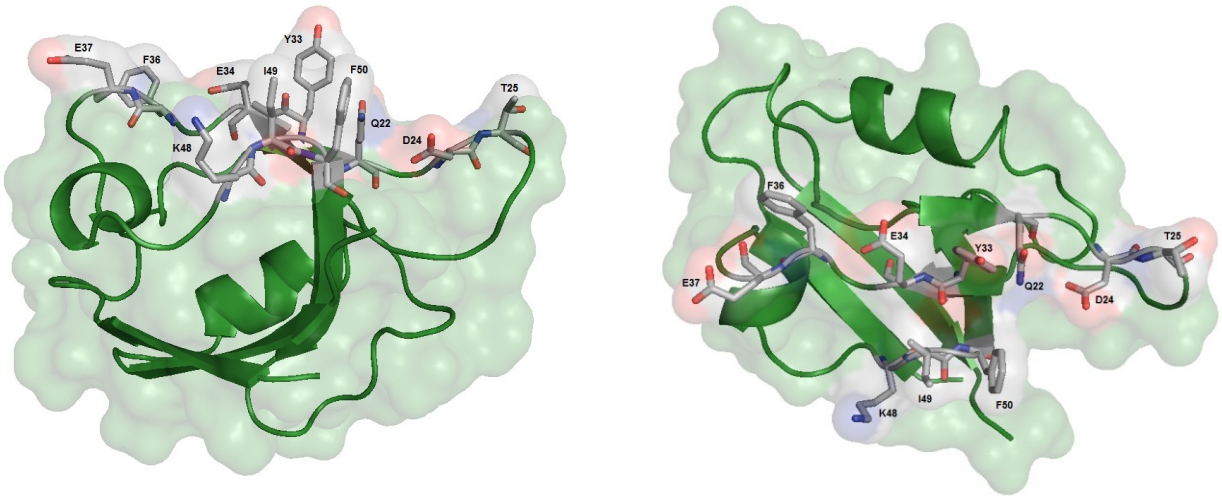


Figure 8 | **Structure of wtPDZ domain.** A | Structure of the wtPDZ domain with highlighted amino acids in the predicted binding site to FOXO4-DBD which were selected for mutagenesis (C – gray, N – blue, O – red). B | The rotated PDZ structure fig (A) by 90 ° around the x-axis toward the reader. Visualization was created in PyMOL software based on PDB ID 3VQF.



A B

Figure 9 | **Mutant structure of the PDZ domain with all ten predicted mutations.** A | Structure of the 10xM PDZ domain with highlighted amino acids in the predicted binding site to FOXO4-DBD which were selected for mutagenesis (C – gray, N – blue, O – red). B | The rotated PDZ structure fig (A) by 90 ° around the x-axis toward the reader. Visualization was created in PyMOL software based on PDB ID 3VQF.

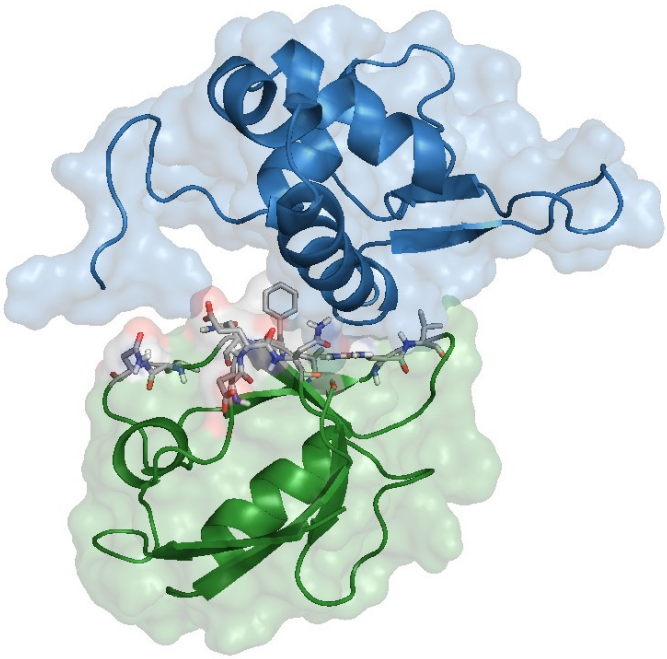


Figure 10 | **PDZ-FOXO4-DBD**. One of the *in silico* predicted docking wtPDZ domain (green) with FOXO4-DBD (blue). Amino acids predicted for binding optimization are highlighted. Visualization was created in PyMOL software.

2 Aims of the project

The main aim of this thesis was to experimentally verify *in silico* predicted protein binders to the FOXO4 transcription factor which plays a double-edged sword role in cancer, and to analyze the PDZ/FOXO4 binding interface in order to help improve the algorithm for *in silico* PPI prediction.

To address this long-term goal, the project plan included the following subtasks:

1. Site directed mutagenesis of the predicted PDZ mutant variants and cloning PDZ mutant variants into EGFP-expression vector.
2. Cloning DBD-FOXO4 into expression vector.
3. Protein expression and purification
4. Measurement of protein stability and PDZ/FOXO4 binding affinity
5. PDZ/FOXO4 binding interface analysis by HDX-MS and NMR

Due to an unsuccessful GAUK grant application in the first year and the lack of institutional funding for my master thesis I was not able to finish the promising evaluation of the binding affinity for all predicted PDZ mutant variants and follow up with the structural analysis. However, I took this failure as an opportunity and applied for an Erasmus+ internship at the Bioinformatics Research Centre in Aarhus, Denmark (supervisor: Palle Villesen). The main task of my internship project was to analyze the expression levels of FOXO family members in primary BCa tumors and tumor infiltrating leukocytes from bulk tumor RNA-Seq expression data with respect to disease progression. Additional tasks were to explore the relationship between cell types and FOXO expression levels in individual BCa patients and between metabolism pathways and FOXO expression levels. The expected output was a dataset of immune and FOXO signatures correlating with BCa disease status and progression prediction

In addition, I decided to apply for the GAUK grant for the second time which included preliminary results presented in this thesis. This application was successful and therefore the planned project will proceed in the following year.

3 Material and Methods

3.1 Buffers and Solutions

Buffer name	Reagents	Concentration	
BufferA_IEC	Tris-HCl (Jersey Lab Supply, USA) NaCl (Lachema, CR)	50 mM 20 mM	pH=7,4
BufferB_IEC	Tris-HCl NaCl	50 mM 1 M	pH=7,4
Lysis_buffer	Tris-HCl Sucrose (Lachema) NaN_3 (Sigma-Aldrich, USA) PMSF* (Sigma-Aldrich) Leupeptin* (Sigma-Aldrich)	50 mM 25% (w/v) 1 mM 1 mM 1 uM	pH=7,4
MST_buffer	Tris-HCl NaCl Tween-20 (Sigma-Aldrich)	50 mM 150 mM 0.05% (v/v)	pH=7,4
BufferA_IMAC	Sodium phosphate NaCl Imidazole	20 mM 500 mM 15 mM	pH=7,4
BufferB_IMAC	Sodium phosphate (Sigma-Aldrich) NaCl Imidazole (Sigma-Aldrich)	20 mM 500 mM 400 mM	pH=7,4
BufferB0_IMAC	Sodium phosphate NaCl	20 mM 500 mM	pH=7,4
Buffer_GF	Tris-HCl NaCl	50 mM 150 mM	pH=7,4

*added just before use

LB medium: 1% (w/v) Tryptone (Oxoid, USA), 0.5% (w/v) Yeast extract (Oxoid), 1% (w/v) NaCl, pH 7.4

LB agar: 1.5% (w/v) Agar (Oxoid) in LB medium

Destaining_solution (SDS-PAGE): 35% (v/v) Ethanol (Lachema), 10% (v/v) Acetic acid (Lachema)

Staining_solution (SDS–PAGE): 45% (v/v) Methanol (Lachema), 10% (v/v) Acetic acid,

0.25% (w/v) Coomassie Brilliant Blue R-250 (Serva, Germany)

Electrode_buffer (SDS–PAGE): 10mM Tris-HCl, 250mM Glycine (Sigma Aldrich), 0.1% (w/v) SDS (Jersey Lab Supply, USA), pH 8.3

2x_reducing_SDS_sample_buffer: 100mM Tris-HCl, 20% (v/v) Glycerol (Lachema), 4% (w/v) SDS, 0.2% (w/v) Bromophenol blue (Serva), 200mM DTT (Sigma Aldrich), pH 6,8

3.2 Molecular cloning

3.2.1 Cloning EGFP into pHis2 plasmid

The EGFP DNA sequence was amplified by polymerase chain reaction (PCR) from a vector containing the EGFP sequence using primers for amplifying EGFP with the following reaction setup:

Q5 high-fidelity 2X master mix (NEB) - 10 µl

Template DNA [5 ng/ µl] – 1 µl

10 µM Forward primer – 1 µl

10 µM Reverse primer – 1 µl

H_2O to the final volume of 20 µl – 7 µl

Program: 98 °C - 1 min – initial denaturation (98 °C – 8 s, 66 °C – 20 s, 72 °C – 20 s) – 30x, 72 °C – 2 min. The PCR product was checked on an agarose gel and then purified with QIAquick PCR Purification Kit (Qiagen, Germany) according to the manufacturer's protocol. Purified EGFP DNA (26 µl) was digested with PagI (0.5 µl) and NotI (0.5 µl) (both Thermo Scientific, USA) restriction enzymes (1h, 37°C) and purified with QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's protocol. The PagI/NotI-digested EGFP DNA (6.5 µl) was ligated into the pHis2 plasmid digested by NcoI and NotI (2 µl) with T4 DNA ligase (NEB, USA) (0.5 µl) in T4 DNA ligase 10x reaction buffer (NEB) (3 µl) for 1 h at room temperature. In addition, a negative control without the insert was included.

Three µl of ligation reaction was used to transform 60 µl of chemically competent DH5-Alpha *E.coli* bacteria (made in house) – transformation reaction: bacteria were thawed on ice, bacteria with ligation reaction were incubated for 20 min on ice, then heat shock - 42 °C - 42 s, 1 min on ice, 1 ml of LB media was added and bacteria were incubated for 1 h at 37°C. After that the cells were spun down at 21 000x g for 1 min and the supernatant was discarded. The suspension of bacteria was plated on Amp+ plates and grown overnight at 37°C. In addition, a negative control without the insert was included.

Four colonies were pre-screened with colony PCR reaction: template from picked colony, 1 µl of each forward and reverse 10 µM primer, 7.5 µl of Taq 2x master mix (NEB), 5.5 µl of H_2O, total volume 15 µl; program: 95°C - 5 min – (95 °C – 30 s, 58 °C – 30 s, 72°C – 60 s) – 30x, 72 °C – 5 min. The PCR products were loaded on a 1.5% agarose gel with GeneRuler 1 kb plus DNA ladder (Thermo Scientific). Positive colonies were inoculated into 4 ml LB media with 150 µg/ml Amp and cultivated 12-16 h at 37°C with

shaking at 220 RPM. Plasmids were isolated with QIAprep Spin Miniprep kit (Qiagen) according to the manufacturer's protocol and their concentrations were measured by spectrophotometer DeNovix DS-11 (DeNovix, USA). The plasmids were sequenced in GATC Biotech AG in Germany and sequences were checked with ApE software (version 2.0.47 for Windows).

3.2.2 Cloning wtPDZ and PDZ 10xM into pHis2-EGFP plasmid

DNA sequences of wtPDZ, mutant variant of PDZ domain with ten mutations (PDZ 10xM) and primers with restriction sites (highlighted green - shown below) were ordered from GenScript (New Jersey, USA).

PDZ NcoI_fwd NNNC'CATGGAAGCTTCCATGTGATTCTG

PDZ NotI_rev NNNGC'GGCCGCTATTATTGACGGCTCACAAC

NcoI and NotI restriction sites were introduced into the PDZ sequences by PCR with the following reaction setup:

Q5 high-fidelity 2X master mix (NEB) - 10 µl

Template wtPDZ/PDZ 10xM DNA – 0.3 µl

10 µM Forward primer – 1 µl

10 µM Reverse primer – 1 µl

H_2O to the final volume of 20 µl – 7.7 µl

Program: 98 °C - 30 s – initial denaturation (98 °C – 8 s, 55 °C – 20 s, 72 °C – 10 s) – 30x, 72 °C – 2 min. The PCR products were checked on a 1.5% agarose gel with GeneRuler 1 kb plus DNA ladder (Thermo Scientific) and then purified with QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's protocol. Purified wtPDZ/PDZ 10xM DNA (26 µl) was digested with NcoI (0.5 µl) and NotI (0.5 µl) (both Thermo Scientific) restriction enzymes (1h, 37°C) and purified with QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's protocol. The NcoI/NotI-digested wtPDZ/PDZ 10xM DNA (6.5 µl) was ligated into the pHis2-EGFP plasmid digested by PagI and NotI (2 µl) with T4 DNA ligase (NEB) (0.5 µl) in T4 DNA ligase 10x reaction buffer (NEB) (3 µl) for 1 h at room temperature. Next, plasmid transformation, colony PCR and plasmid isolation was performed according to the procedure mentioned in "3.2.1 Cloning EGFP into pHis2 plasmid" (above). The plasmids were sequenced in GATC Biotech AG in Germany and the sequences were checked with ApE software (version 2.0.47 for Windows).

3.3 Whole plasmid site-directed mutagenesis

The mutations were introduced at the 5' end of the non-phosphorylated reverse primers. The forward primers were phosphorylated and both primers were designed longer than 15 bases with Tm above 60°C (see primer sequences in Supplementary information). Site-directed mutagenesis by PCR was performed with the following reaction setup:

Q5 high-fidelity 2X master mix (NEB) - 10 µl

Template pHis2-wtPDZ-EGFP DNA – 0.3 µl

10 µM Forward primer – 1 µl

10 µM Reverse primer – 1 µl

H_2O to the final volume of 20 µl – 7 µl

Program: 98 °C - 1 min – initial denaturation (98 °C – 8 s, 64 °C – 20 s, 72 °C – 4 min 9 s) – 30x, 72 °C – 5 min. The PCR products were checked on a 1.5% agarose gel with GeneRuler 1 kb plus DNA ladder (Thermo Scientific). Positive samples from PCR (4 µl) were ligated with T4 DNA ligase (NEB) (0.5 µl) in T4 DNA ligase 10x reaction buffer (NEB) (3 µl) with 1.5 µl of PEG4000 (because of blunt-end ligation) and 7.5 µl of H_2O to the final volume of 15 µl for 1 h at room temperature. Then transformation and isolation were performed according to the procedure mentioned in "3.2.1 Cloning EGFP into pHis2 plasmid" (above). The plasmids were sequenced in GATC Biotech AG in Germany and the sequences were checked with ApE software (version 2.0.47 for Windows).

Several PCR mutagenesis reactions yielded no product and some sequencing results showed wtPDZ instead of the mutant variant. To troubleshoot PCR mutagenesis, annealing temperature was decreased and template wtPDZ DNA was 1000 times diluted. This helped overcome the issue with negative sequencing results, but it still did not work for every PCR mutagenesis. Thus gradient PCR was conducted to determine the optimum annealing temperature with the temperature gradient ranging from 54 °C to 63 °C.

3.4 Protein expression

3.4.1 FOXO4 protein expression

A half µl of pET-15b plasmid containing protein-coding DNA for the FOXO4-DBD(82-207) protein with N-terminal 6x-His fusion tag (Boura et al., 2007) was used to transform 40 µl of chemically competent bacteria *E.coli* BL-21(DE3) Gold (Stratagene, USA) – transformation reaction: bacteria were thawed on ice, bacteria with the expression

plasmid were incubated for 10 min on ice, then heat shock - 42 °C - 42 s, 1 min on ice, 400 µl of LB media was added and bacteria were incubated for 5 min at room temperature. Then 5 ml of LB media with 150 µg/ml Amp was added and the bacteria were grown overnight at 37°C with shaking at 220 RPM.

Protein production was carried out by adding 1 ml of the starter culture from the previous step to 0.5 l of LB media with 150 µg/ml Amp and bacteria were grown at 37°C with shaking at 220 RPM until the optical density of the sample measured at wavelength of 600 nm reached 0.6-0.8 (measured by spectrophotometer DeNovix DS-11 (DeNovix)). Expression of the recombinant protein FOXO4-DBD was induced by addition of IPTG (Isopropyl β-D-1-thiogalactopyranoside) (Sigma-Aldrich) to a final concentration of 1 mM (the protocol for FOXO4 production is well established in the lab and therefore I did not optimize IPTG concentration). The bacteria were grown for 12 h at 30 °C with shaking at 220 RPM. After that they were harvested by centrifugation at 6 000x g for 10 min at 4 °C. The cell pellet was washed by lysis_buffer and resuspended in 20 ml/l of the same buffer. The cell suspension was supplemented with lysozyme (5 µg/ml), DNase I (300 U), RNase (300 U), 400 µl of 1M MgCl2 and was frozen to -80 °C, thawed and sonicated for 4 min on ice. The sonicated cell extract was centrifuged at 40 000x g for 20 min at 4 °C and the supernatant was collected.

3.4.2 PDZ and EGFP protein expression

A half µl of plasmid pHis2 containing protein-coding DNA for wtPDZ/PDZ M3/PDZ M5/PDZ M10/PDZ 10xM (all fusion proteins with EGFP) or EGFP with an N-terminal 6-His fusion tag was used to transform 40 µl of chemically competent bacteria *E.coli* BL-21(DE3) Gold (Stratagene, USA) – transformation reaction: according to the procedure mentioned in “3.4.1 FOXO4 protein expression” (above).

To find the optimal IPTG concentration for protein expression induction one ml of LB medium with 150 µg/ml Amp was added into six microtubes. Each microtube was inoculated with 10 µl of the overnight grown bacterial culture. Then bacteria were grown at 37 °C with shaking at 220 RPM for 1 hour. After that IPTG (Sigma-Aldrich) was added into each microtube in the following concentrations: 1 mM, 0.5 mM, 0.2 mM, 0.1 mM, 0.05 mM and 0 mM. Bacteria were incubated at 37°C with shaking at 220 RPM for 2.5 hour and then centrifuged at 5000x g for 10 min at room temperature. The bacterial pellets were resuspended in 100 µl of 1x SDS reduction buffer, boiled for 5 min and analyzed by SDS-PAGE.

Protein production was conducted according to the procedure in “3.4.1 FOXO4 protein expression” (above) with the optimized IPTG concentration at 0.1 mM.

3.5 SDS-PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed with a 15% polyacrylamide separating gel and 5% polyacrylamide stacking gel. Gels were prepared according to the table below:

	15% polyacrylamide separating gel [ml]	5% polyacrylamide stacking gel [ml]
water	1.1	0.68
30% acrylamide mix	2.5	0.17
1.5M Tris (pH 8.8)	1.3	x
1.0M Tris (pH 6.8)	x	0.13
10% Sodium dodecyl sulfate (SDS)	0.05	0.01
10% ammonium persulfate (APS)	0.05	0.01
Tetramethylethylenediamine (TEMED)	0.002	0.001
Total	5	1

Fifteen µl of sample with 5 µl of 2x reducing SDS sample buffer were boiled for 5 min and then centrifuged at 10 000x g for 10 min at room temperature. The electrophoresis apparatus was filled with electrode_buffer and the centrifuged samples were loaded on the gel. Electrophoresis was run at a constant current 15 mA/gel for stacking gel and 25 mA/gel for separating gel. After electrophoresis, the gels were stained with staining_solution for 15 min and the non-specifically stained background was washed out with destaining_solution.

3.6 Protein purification

3.6.1 Immobilized metal affinity chromatography

Immobilized metal ion affinity chromatography (IMAC) is usually used as a first step in protein purification. It separates proteins on the basis of reversible interactions between an amino acid tag (e.g. 6xHis-tag) and immobilized metal ions. For example, TALON superflow resin (Clontech, USA) consist of a tetradentate chelator charged with cobalt that has an affinity and a specificity for a polyhistidine-tag. Elution of the bound protein is performed with a buffer containing a high concentration of imidazole that acts as a competitive ligand for the metal ions (reviewed in (Cheung et al., 2012)).

A column packed with TALON superflow resin (Clontech) was equilibrated with BufferB0_IMAC and washed with BufferA_IMAC. The collected supernatant from protein production was loaded on the column at a flow rate of 1ml/min. The proteins bound to the column were washed with the BufferA_IMAC and then eluted with BufferB_IMAC. The eluate was analyzed by SDS-PAGE.

3.6.2 Size exclusion chromatography

Size exclusion chromatography (SEC) separates molecules based on their size and is usually used as a second step in protein purification when higher purity is required. SEC is also a convenient method for removing possible aggregates that would interfere with downstream binding assays.

The column EnRich Sec70 10/300 24 ml (Bio-Rad, USA) was connected to the pump and equilibrated with buffer_GF at a flow rate of 0.8 ml/min for 35 min. The concentrated eluate from IMAC was loaded on the column at a constant flow rate 0.5 ml/min. Fractions of 1 ml were collected after the absorbance of the solution eluted from the column (measured at wavelength of 280 nm) exceeded 0.1 threshold. The fractions were analyzed by SDS-PAGE and protein concentrations measured by spectrophotometer DeNovix DS-11 (DeNovix). Purified proteins were stored at -80 °C.

3.6.3 Ion exchange chromatography

Ion exchange chromatography (IEC) was performed to remove the possible contamination of the FOXO4 protein sample with PDZ-EGFP protein. It separates molecules on the basis of their net surface charge and elution is performed by increasing ionic strength that can be controlled by using different forms of salt gradients.

The column Mono S HR 5/5 (GE Healthcare Life Sciences, USA) was connected to the pump and was washed with the BufferA_IEC at a flow rate of 1 ml/min until the conductivity was stable. After that the sample was loaded at a flow rate of 1 ml/min. The elution was performed with a linear salt (NaCl) gradient from 20 mM to 1 M at a flow rate of 1ml/min for 50 min. Fractions of 1 ml were collected after the absorbance of the solution eluted from the column (measured at 280 nm) exceeded 0.1 threshold. Samples were desalted by Amicon Ultra-15 10K centrifugal filter device (Merck Millipore, USA) according to the manufacturer’s protocol. Purified proteins were stored at -80 °C.

3.7 Spectrofluorimetry

The synchronous fluorescence spectrum of the FOXO4 sample [150 µM] was measured in the range from 320 nm to 800 nm with a 15 nm constant difference between excitation and emission wavelength. The synchronous spectrum of the wtPDZ-EGFP [0.1 µM] was measured under the same conditions for comparison. The fluorescence spectra were measured by Spectrofluorometer FluoroMax 3 (Jobin Yvon)

3.8 Nano differential scanning fluorimetry

Prometheus NT.48 (NanoTemper Technologies GmbH, Germany) induces thermal unfolding of proteins and detects changes of the intrinsic fluorescence of the amino acid tryptophan (Trp), which is usually located within the hydrophobic core of a protein. When a protein unfolds the Trp gets exposed to the different environment. This leads to a change in its fluorescence that is detected by highly sensitive UV-detectors. By detecting the changes in the Trp fluorescence, protein quality, aggregation, correct protein folding and the melting temperature (T_m) can be determined in a dye-free approach (Maschberger and Breitsprecher, 2015). Amino acids tyrosine and phenylalanine contribute to the protein intrinsic fluorescence significantly less than Trp, but it was shown that they can be exploited for the measurement with the Prometheus NT.48 when a measured protein sequence does not contain any Trp (Misetic, 2016).

NanoDSF grade high sensitivity capillaries (NanoTemper Technologies GmbH) were filled with 10 µl of sample per capillary. Capillaries were dipped into the samples, filled automatically by capillary forces, placed into the machine and heated at a rate of 1 °C/min, from 20 °C to 95 °C. Fluorescence at 350 nm and 330 nm was collected at a rate of 10 data points per minute. The data were analyzed with PR.control software (version 1.1) from NanoTemper Technologies GmbH. Unfolding transition midpoints were automatically calculated by the software from the first derivative of fluorescence ratio (F350/F330).

3.9 Microscale thermophoresis

The interaction of molecules alters their size, charge or hydration shell which causes the changes of the mobility of molecules in microscopic temperature gradients. These changes can be detected by MicroScale Thermophoresis (MST) that allows detection of biomolecular interactions from ion binding up to interactions of large complexes (e.g. ribosomes) under close-to-native conditions. The instrument Monolith

NT.115 pico (NanoTemper Technologies GmbH) analyzes thermophoresis by measuring fluorescence and therefore one of the binding partners must be fluorescently labeled.

During the binding experiment concentration of the fluorescent molecule is kept constant and concentration of the titrant is altered by titration. Binding of a titrated, unlabeled molecule to a fluorescent molecule alters its size, charge or hydration shell. This can be quantified by measuring the change in thermophoresis (directed movement of molecules along a temperature gradient) of the fluorescent molecule induced by infrared laser that is monitored in titration experiments with Monolith NT.115 pico instrument (NanoTemper Technologies GmbH). The dissociation constant K_d is then automatically calculated with Affinity Analysis (MST) software (version 2.0.2) by fitting the binding curve (Figure 11).

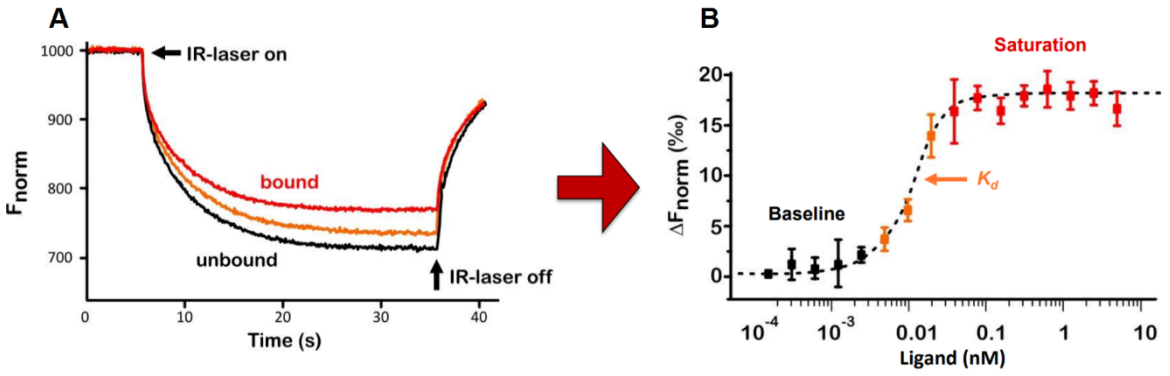


Figure 11 | **Microscale thermophoresis.** A **|** The thermophoresis inside the capillary is initiated with a focused IR-laser (IR-laser on). As the temperature inside the capillary increases the fluorescent molecules move away and the fluorescence decreases. The fluorescent molecules that formed complexes with titrated molecules move away at a slower rate (bound vs unbound). B | Plotted normalized fluorescence signal from each capillary with a different concentration of titrant molecules results in a binding curve that is subsequently used for fitting K_d. Adapted from https://nanotempertech.com/.

Protein concentrations were measured by spectrophotometer DeNovix DS-11 (DeNovix). FOXO4 samples were concentrated by Amicon Ultra-15 10K or Amicon Ultra-0.5 10k centrifugal filter devices (Merck Millipore) according to the manufacturer's protocol. Before the binding experiments two types of capillaries were checked to avoid possible protein adsorption to the glass capillary, which could result in experimental artifacts. Monolith NT.115 MST standard treated capillaries and Monolith NT.115 MST premium coated capillaries (both NanoTemper Technologies GmbH) were filled with fluorescently labeled PDZ [3.2 nM] (in MST_buffer) and their fluorescence was measured by Monolith NT.115 pico instrument (NanoTemper Technologies GmbH). On the basis of these measurements I chose premium coated capillaries that showed no sticking

compared to slight sticking when using standard treated capillaries. Buffer composition (NaCl 150 mM, Tris-HCl 50 mM, pH=7.4) was optimized with 0.05% Tween-20 to reduce noise as it is recommended in the manufacturer's protocol.

Both PDZ and FOXO4 were centrifuged at 20 000x g for 10 min at room temperature to remove protein aggregates. Then 5 µl of MST_buffer was added to 9-15 PCR tubes and 10 µl of FOXO4 ligand [556 µM] was pipetted to PCR tube one. A serial dilution of unlabeled FOXO4 was prepared: 5 µl from PCR tube one was transferred to PCR tube two, from two to three - repeated for all PCR tubes. After that 5 µl of fluorescently labeled PDZ [3.2 nM] was added to each PCR tube of the 2-fold dilution series and mixed well by pipetting up and down. The binding reaction was incubated for 10 min at room temperature. Monolith NT.115 MST premium coated capillaries (NanoTemper Technologies GmbH) were dipped into the samples, filled automatically by capillary forces, placed into the machine and MST analysis was started with MST power 40 % and 60 %. As a negative control EGFP protein was used instead of PDZ-EGFP fusion protein. Data were analyzed with Affinity Analysis (MST) software (version 2.0.2).

3.10 Bioinformatic analysis of bladder cancer RNA-Seq data

3.10.1 BCa RNA-Seq data

This thesis includes BCa RNA-Seq data (only first diagnosed tumors) from previous studies (Hedegaard et al., 2016) (N=476) and (Nordentoft et al., 2014) (N=29) that are deposited in the EGA (https://www.ebi.ac.uk/ega/) under accession numbers EGAS00001001236 and EGAS00001001686 respectively. In addition, unpublished RNA-Seq data from 21 patients (first diagnosed tumors) are included in this thesis. These samples were prepared and sequenced following the same procedure as in (Nordentoft et al., 2014).

Samples (N=476) from (Hedegaard et al., 2016) were collected from year 2008 to 2012 in multiple European hospitals. Overall the progression rate was low – only 31 samples (~ 7 %) were progressive tumors. Thus I decided to collect mores samples to increase the number of progressive tumors. I included unpublished samples (N=21) and samples (N=29) from (Nordentoft et al., 2014) which were collected in Aarhus University Hospital. These datasets contained 26 progressive tumors. All patients were followed according to the national guidelines.

Altogether I collected 510 NMIBC (Cis, Ta and T1) RNA-Seq samples and 16 MIBC (T2+) RNA-Seq samples. Within NMIBC 453 were non-progressive tumor samples and

57 progressive tumor samples (Figure 12). For each sample clinical information about patient age, patient sex, patient progression free survival (PFS), tumor stage, tumor grade, cancer progression and intravesical BCG (Bacillus Calmette–Guérin) treatment were available.

Here I worked with gene FPKM (Fragments Per Kilobase of exon per Million fragments mapped) matrices that were all generated with the Tuxedo Suite tools (Trapnell et al., 2012) using the same procedure.

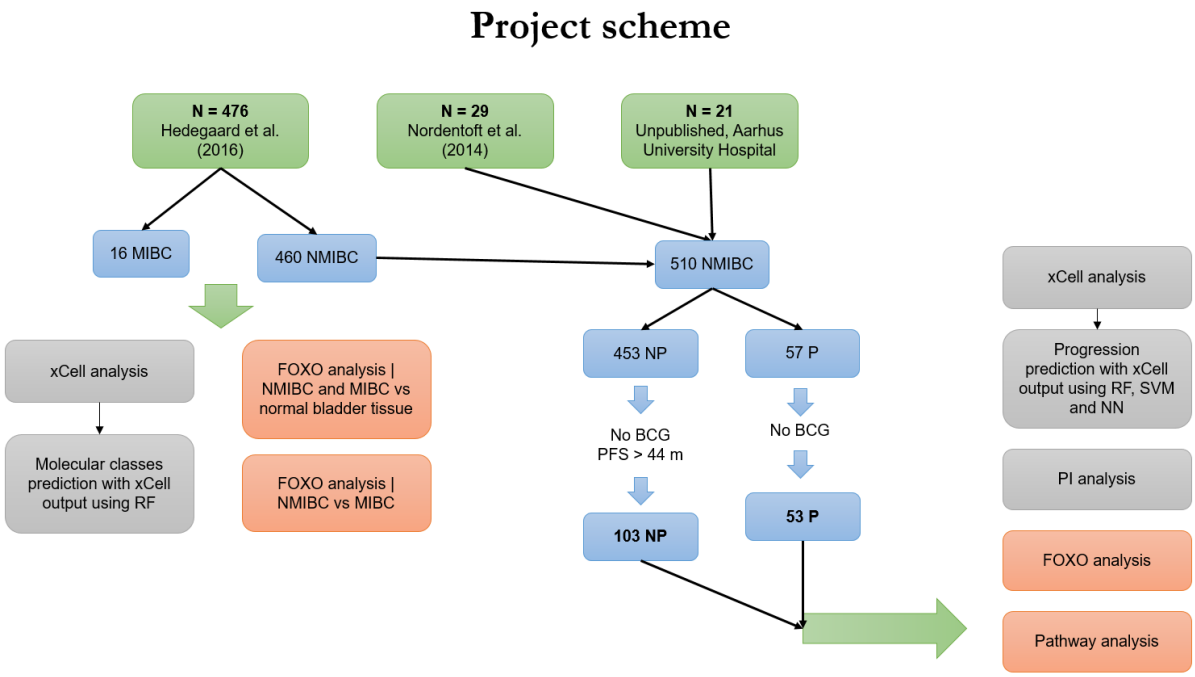


Figure 12 | **Project scheme.** Altogether 526 transcriptomic samples were collected from first diagnosed bladder tumors (green boxes). 476 samples from (Hedegaard et al., 2016) were used for FOXO analysis (red) and xCell analysis (gray). Then all NMIBC were merged and filtered. 103 non-progressive (NP; PFS > 44 m, no BCG) and 53 progressive (P; no BCG) BCa samples were used for FOXO and pathway analysis (red), and xCell and para-inflammation (PI) analysis (gray).

3.10.2 BCa TME analysis by xCell method

xCell method (Aran et al., 2017) is a gene signature based method for tumor digital dissection. xCell gene signature set enables to infer 64 cell types and estimate tumor purity. Aran et al. (2017) found that tumor purity negatively correlates with the microenvironment score, which was the average score of 47 immune and stroma cell types.

The gene FPKM matrices were read into the R computing environment (version 3.3.3 for Windows) and merged into one file with genes in rows and samples in columns.

This file was then uploaded to xCell webtool (available at: http://xcell.ucsf.edu/) and the analysis was performed with xCell gene signature set (N=64). The result from the analysis was read into the R computing environment (version 3.3.3 for Windows) and all the following analyses were conducted within this environment. Samples from NMIBC patients were split into two groups. The first group contained only non-progressive tumors (NP) and the second only progressive tumors (P). Tumors were classified as progressive if NMIBC (Cis, Ta and T1 stage) progressed to MIBC (T2+). Then the file containing clinical information about samples was joined to the xCell file and the clinical information was examined. The highest PFS in P group was 44 months (average = 11 m, median = 6 m) and in NP group 75 months (average = 36 m, median = 35 m). Since not all patients were followed for five years due to patient death or other reasons there could be tumors within NP group that would eventually progress. Thus, preliminary analysis with PFS cutoffs set to 12, 24 and 44 months were performed. Indeed, the significance increased with higher PFS cutoff. For further analysis, the highest cutoff (PFS>44 months) was selected to reduce the noise from misclassified tumors to a minimum. Data distribution was visually inspected with qqPlot function from car package (version 2.1-5) though normality assumption for parametric tests is not vital for large numbers of samples (N>50). Differences in cell types between NP group and P group were tested with t-test with unequal variances using t.test function. Data were visualized with boxplot function and with functions from beanplot (version 1.2) package.

3.10.3 Machine learning analysis

3.10.3.1 Random Forest

The Random Forest (RF) machine learning algorithm was performed in the R computing environment (version 3.3.3 for Windows) using randomForest package (version 4.6-12). Altogether 103 NP (PFS > 44m, no BCG) and 53 P (no BCG) samples were randomly split into the training and the test set in the ratio 8:2. Subsequently, RF with all cell type variables from the xCell output was run to rank their importance in the prediction model with mean decrease accuracy (MDA). Then RF was run again with the cell showing the highest MDA value and each additional run one more cell type having lower MDA was added and the model performance was estimated. After that the model with the highest accuracy was selected and its performance was tested on the test set.

Then patient age and/or tumour grade were added into the model to improve the performance and the same procedure was repeated. Due to a different grading system between dataset 1 (low/high grade; WHO 2004) and dataset 2 (grade 1, 2 or 3; WHO

1973) only grade 1 and grade 3 samples from dataset 2 were taken and denoted as low grade and high grade respectively. There were no PUNLMP tumors. RF with cells, age and grade, cells and age and only with cells were conducted.

For BCa molecular classes prediction 476 samples from (Hedegaard et al., 2016) were used without splitting them into the training and the test set. All Random Forest algorithms were performed with a 1000 trees to grow (= ntree parameter). Importance and proximity parameters were set to TRUE. Number of randomly sampled variables available for splitting at each tree node (= mtry parameter) was set to sqrt(number of variables in the dataset). RF results were visualized with multidimensional scaling (MDS) plot using MDSplot function. Confusion matrices were created with confusionMatrix function. RF prediction models were tested on the test set with predict function. Matthews Correlation Coefficient (Matthews (1975) cited in Baldi et al. (2000)) was calculated from the following equation: MCC = ((TruePositive(TP)* TrueNegative(TN)) - (FalsePositive(FP)* FalseNegative(FN))) / sqrt((TP+FP) * (TP+FN) * (TN+FP) * (TN+FN)). All repetitive tasks were automatized using loops or functions from the Apply family of functions (e.g. apply, lapply, sapply).

3.10.3.2 Support Vector Machine

The Support Vector Machine (SVM) machine learning algorithm was performed in the R computing environment (version 3.3.3 for Windows) using e1071 package (version 1.6-8). The progression prediction models were trained with linear kernel using tune.svm function on the same training sets that were used for RF. Then tune function was run to choose the best cost parameter and 10x cross-validation to estimate the model performance and subsequently the best model was tested on the test set using the predict function.

3.10.3.3 Neural Network

The Neural Network (NN) machine learning algorithm was performed in the R computing environment (version 3.3.3 for Windows) using neuralnet package (version 1.33). The training sets that were used for RF were randomly split into the new training sets and the validation sets in the ratio 8:2. The NN was trained on the new training sets using neuralnet function with 1:64 number of hidden neurons and the models' performances were estimated on the validation sets. Subsequently the models with the highest accuracy were chosen and tested on the test sets using the compute function.

3.10.4 PI score

The PI scores were calculated with R programming language code ('calculatePIscore' function) and objects, that were acquired from Dvir Aran,Ph.D (Aran et al., 2016) via email correspondence, in the R computing environment (version 3.3.3 for Windows). 'CalculatePIscore' function is a ssGSEA (single-sample Gene Set Enrichment Analysis) method that calculates ssGSEA score based on a signature of 40 genes (see gene signature in Supplementary information). Since the differences between PI scores could be explained only by differences in purity levels of the samples, an adjustment procedure is included in 'calculatePIscore' to remove inflammatory gene expression originating from immune infiltrations. The PI scores were calculated for 103 NP (PFS > 44m, no BCG) and 53 P (no BCG) samples using 'calculatePIscore' function and the results were visualized with beanplot function. Differences in the PI scores between NP group and P group were tested with t-test with unequal variances using the t.test function.

(Scripts and objects for calculating the PI score are available at email address: tausp@natur.cuni.cz)

3.10.5 FOXO analysis

Exploratory analysis, statistical tests and data visualization were performed within the R computing environment (version 3.3.3 for Windows) using functions already mentioned above. Data were manipulated and filtered with functions from dplyr package (version 0.6.0).

3.10.6 Pathway analysis

Whole genome enrichment analysis of 103 NP (PFS > 44m, no BCG) and 53 P (no BCG) samples was conducted using Gene set enrichment analysis (GSEA) (version 2.2.4) and results were visualized using Cytoscape (version 3.4.0). Two files were required for GSEA – rank file (.rnk extension) and pathway database (.gmt extension). Pathway database ("Human_GOBP_AllPathways_no_GO_iea_ May_01_2017_symbol.gmt") was downloaded from http://baderlab.org/GeneSets. To generate a rank file for GSEA a score (sign(mean(NP)-mean(P))*-log10(pvalue)) was calculated from Welch Two Sample t-test differential expression results. The sign of the difference of the means indicates if the gene has an expression which is higher in NP (+ sign) or in P (- sign). The following parameters for GSEA were specified - No. of permutations = 2000, collapse dataset to gene symbols = false, enrichment statistics = weighted, max size = 500 and min size = 10. GSEA performs a gene-set enrichment analysis using a modified Kolmogorov-Smirnov statistic.

The output result consists of summary tables displaying enrichment statistics for each gene-set (pathway) that has been tested. For the pathway visualization with Cytoscape the following files were needed – gmt file and rank file (input files for GSEA), expression file (RNA seq data for all samples and all genes), GSEA ouput 1 (GSEA results for NP phenotype) and GSEA output 2 (GSEA results for P phenotype). EnrichmentMap app within Cytoscape was used to create visualization. For the parameters, P-value cutoff was set to 1 (not used as a selection criteria), FDR Q-value to 1E-05 (only gene-sets enriched at a value 1E-05 or less were displayed) and in similarity cutoff Jaccard+Overlap Combined was selected with default values (cutoff=0.375, combined constant=0.5). To auto-annotate the network with cluster labels Autoannotate app within Cytoscape was used (additionally apps WordCloud and ClusterMaker had to be installed). Markov Cluster algorithm and default label algorithm (WordCloud: Adjacent Words) with max four labels per label and five bonus for adjacent word were used within the app. For clarity of visualization clusters with two or less nodes were removed.

3.10.7 Relationships between FOXO expression, TME and PI score in non-progressive and progressive BCa

The correlation between FOXO expression level, cell types in TME and PI score were tested with a Spearman's correlation test using cor.test function. After statistical tests, significant results ($p < .01$) were visually inspected by plot function to check if these results were not caused by a few outliers. All repetitive tasks were automated using loops or functions from the Apply family of functions (e.g. apply, lapply, sapply).

4 Results

4.1 Microscale thermophoresis binding assay

This study was initiated by the *in silico* identification of the key interface residues in the complex of the wtPDZ domain and FOXO4. Consequently, ten mutations in the PDZ domain which resulted in higher interaction affinity between the two molecules were selected. In this part of my diploma thesis I proved that the PDZ domain structure is indeed suitable for rational design and verified the choice of four mutant variants. Here I present the results of the preparation of predicted PDZ mutant variants, protein production and purification and eventually the binding experiment. Additionally, I show the results of my struggle with fluorescence of non-tagged FOXO4-DBD.

4.1.1 Protein preparation for microscale thermophoresis binding assay

From the first round of PCR mutagenesis I obtained only one mutant variant of the PDZ domain out of ten and from the second round with a decreased annealing temperature and lower DNA template concentration another one out of five. This failure prompted me to perform gradient PCR for one mutant variant with a successful output. Altogether I prepared three DNA sequences of mutant PDZ variants – PDZ M3, PDZ M5 and PDZ M10. Even though gradient PCR mutagenesis was a seemingly successful strategy I could not perform more of them because the lack of funding.

I produced and then purified the following proteins by IMAC and GF: wtPDZ (Figure 13), PDZ M3, PDZ M5, PDZ M10, PDZ 10xM (all PDZ were fusion protein with EGFP), EGFP and FOXO4-DBD (Figure 14).

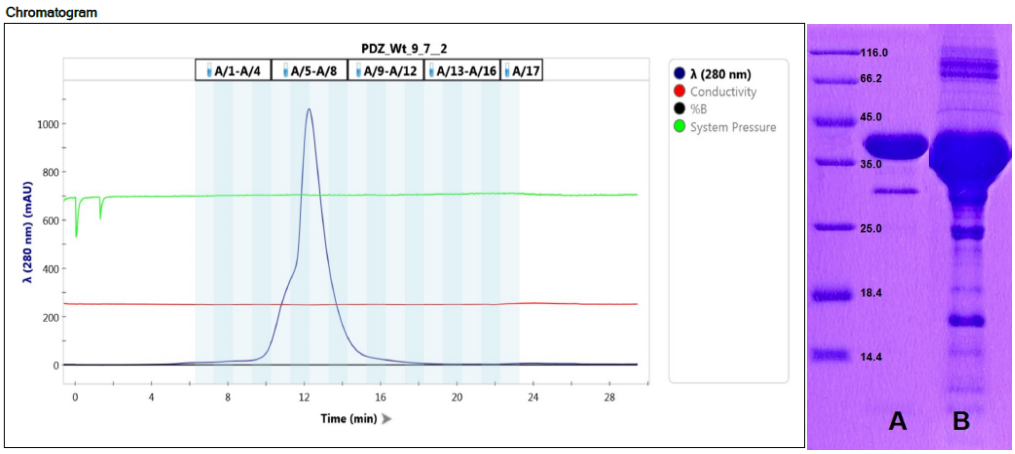


Figure 13 | **SEC chromatogram of wtPDZ purification (on the left) and SDS-PAGE of wtPDZ (on the right).** A | Sample of wtPDZ after purification by SEC. B | Sample of wtPDZ after purification by IMAC and before purification by SEC. (Marker, line 'A' and line 'B' are from one SDS-PAGE gel but were put together in MS Paint for the clarity of visualization)

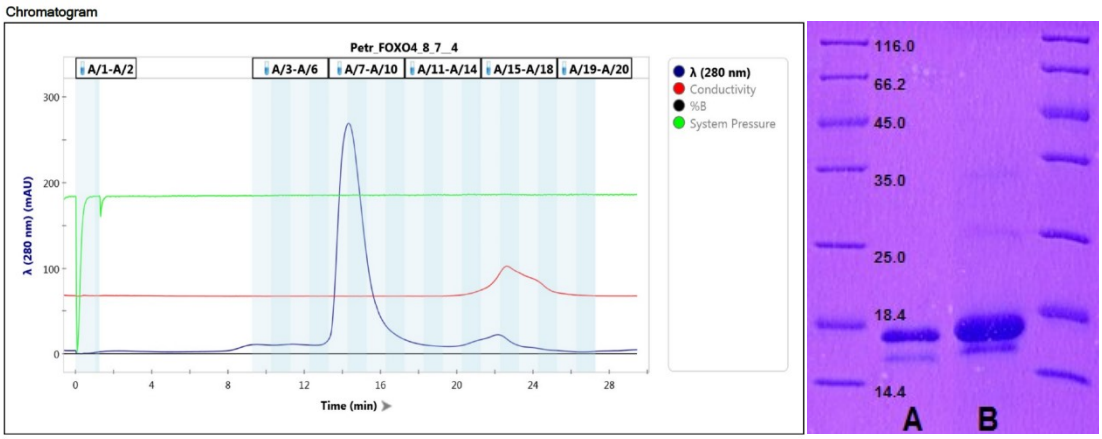


Figure 14 | **SEC chromatogram of FOXO4 purification (on the left) and SDS-PAGE of FOXO4 (on the right).** A | Sample of FOXO4 after purification by SEC. B | Sample of FOXO4 after purification by IMAC and before purification by SEC. (Markers, line 'A' and line 'B' are from one SDS-PAGE gel but were put together in MS Paint for the clarity of visualization)

4.1.2 Issue with fluorescence of non-tagged FOXO4-DBD

When performing the first MST measurement I detected an intrinsic fluorescence of non-tagged FOXO4 in blue and red spectrum. Next, I measured the synchronous fluorescence spectrum of the FOXO4 sample in range from 320 nm to 800 nm with constant difference between excitation and emission wavelength 15 nm and the synchronous spectrum of the wtPDZ-EGFP under the same conditions for comparison (Figure 15). The fluorescent spectrum of wtPDZ-EGFP [0.1 μM] demonstrated a spectrum typical for EGFP protein and the fluorescent spectrum of FOXO4-DBD [150 μM] also showed a peak typical for EGFP but considerably smaller (~17x). This peak interfered with the MST measurement by Monolith NT.115 pico instrument in the blue LED diode spectrum (absorption=460-480, emission=510-530). Additionally, a peak typical for Trp (FOXO4-DBD has 5 Trp) and also a peak in the red spectrum had appeared. The latter mentioned peak corresponded with results of fluorescence measurement by Monolith NT.115 pico instrument in the red LED diode spectrum (Figure 16).

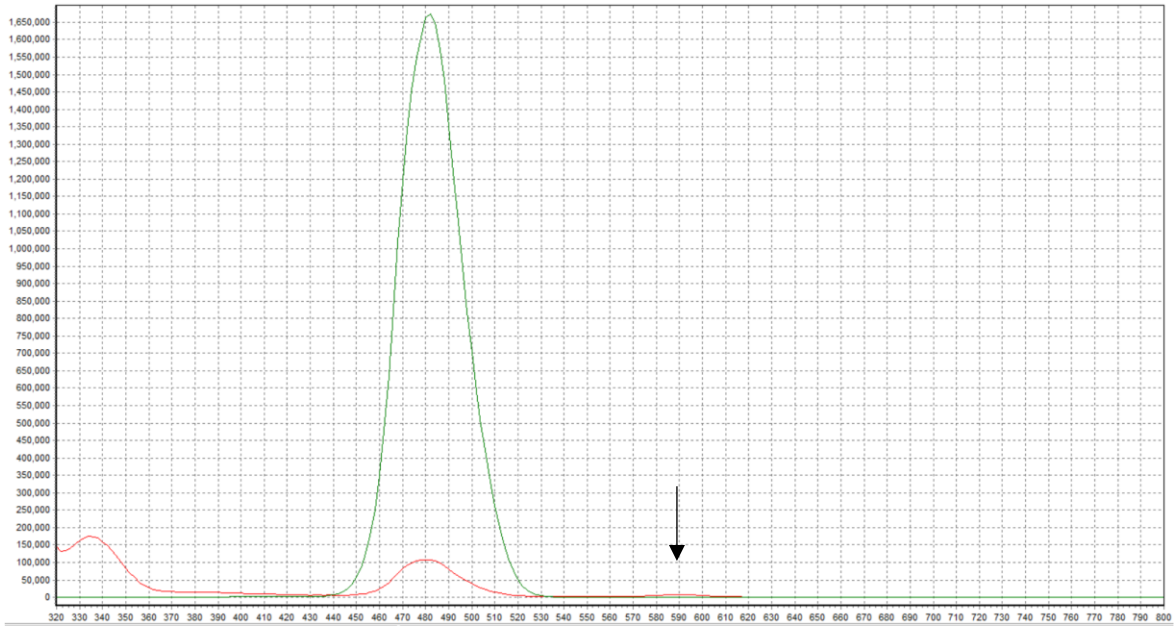
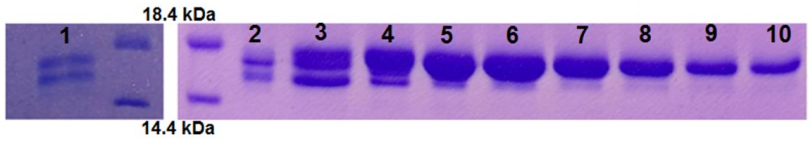


Figure 15 | **Synchronous fluorescence spectrum of 150 µM FOXO4 sample (red) and 0.1 µM wtPDZ-EGFP (green) from 320 nm to 800 nm.** Black arrow shows a small fluorescence peak in 'red' spectrum.

The opposite net surface charge (at pH 7.4) of FOXO4-DBD and PDZ-EGFP fusion proteins prompted me to purify the FOXO4 protein sample with IEC to remove possible contamination with PDZ-EGFP. However, the fluorescence was eliminated only in some IEC fractions. Moreover, this experiment showed that the fluorescence was independent on the protein concentration (Figure 16).



IEC fractions	Concentration [µM]	Fluorescence 'Blue' (absorption = 460-480, emission = 510-530)	Fluorescence 'Red' (absorption = 600-650, emission = 680-690)
1	3.5	660	5250
2	12.5	120	300
3	21.4	100	0
4	37.2	85	0
5	53.4	30	0
6	44.2	40	0
7	25.7	0	0
8	16.3	10	0
9	10.0	0	0
10	6.6	10	0

Figure 16 | **Analysis of FOXO4 fluorescence.** IEC fractions of FOXO4 were analyzed by SDS-PAGE (at the top) and fluorescence was measured in 'Blue' (absorption = 460-480,emission = 510-530) and 'Red' (a. = 600-650,e. = 680-690) spectrum by MST instrument. Fluorescence was independent on the protein concentration.

There were no visible bands on the SDS-PAGE gel in the PDZ-EGFP area. I excised three bands in the range 18.4 – 14.4 kDa and performed MALDI-TOF mass spectrometry with cleavage agent Trypsin/P (mass spectrometry measurement was performed by Daniel Kavan in the Laboratory of Structural Biology and Cell Signalling, IM CAS). All three analyzed bands from the gel were proved to be FOXO4-DBD proteins with the same N-terminus by the comparison of mass spectra.

Next, I performed IEC with a focused salt gradient to separate FOXO4 degradation products with the assumption that these might be the source of the fluorescence. Indeed, I separated degradation products by focused IEC but the fluorescence did not disappear (Figure 17). I did not identify the source of the fluorescence but based on my results I suggest that it was not the contamination by EGFP (after SEC and IEC) nor small molecular weight compounds (after SEC). Eventually I was able to prepare FOXO4 samples without fluorescence or minimal fluorescence that did not interfere with the MST measurement (i.e. random fluorescence changes in capillaries were ≤ ~10 %) by repeating IEC, but the complete procedure was time and cost inefficient.

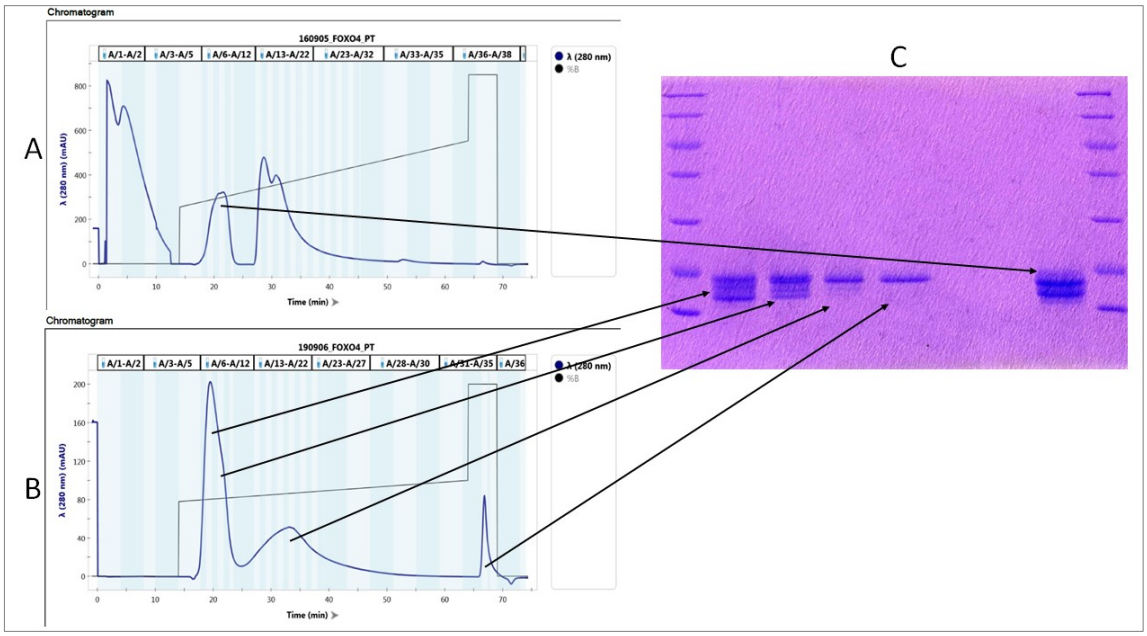


Figure 17 | **IEC with focused salt gradient.** A | Sample after IEC purification was purified again with IEC with focused salt gradient. Two peaks were separated. B | Eluted protein sample from the second peak was collected and purified again with IEC with a more focused salt gradient. C | Degradation products were separated from FOXO4, however the fluorescence was present in this sample. (Peak on (A) chromatogram before the salt gradient was applied is an instrument artefact).

4.1.3 Nano differential scanning fluorimetry and microscale thermophoresis

I performed binding experiments of four PDZ mutant variants and wtPDZ by MST (Monolith NT.115 pico instrument). As a negative control, I used EGFP only. Before the MST binding assays, I measured the quality of all purified PDZ domains and FOXO4-DBD (Figure 18 and 19) with nanoDSF fluorimetry (Prometheus NT.48). These measurements showed that the protein quality was sufficient for binding experiments using MST.

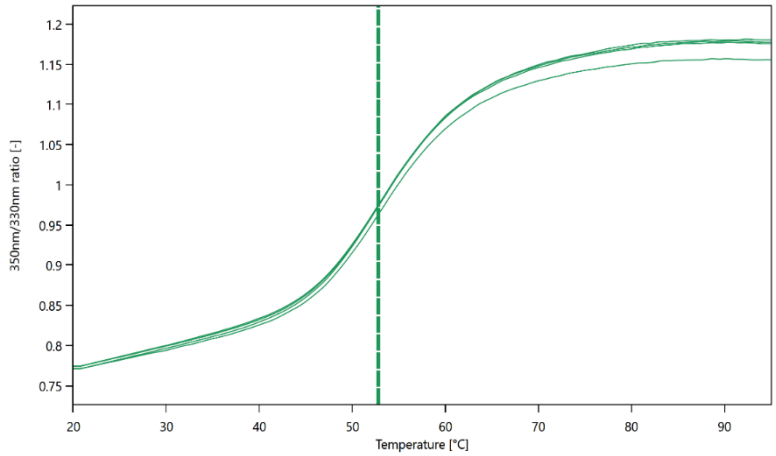


Figure 18 | **FOXO4 thermal unfolding.** Results of qualitative measurement of four FOXO4 technical replicates using Prometheus NT.48. The results show that the quality of the produce protein is sufficient for the next measurements. The estimated denaturation midpoint was ~53 °C.

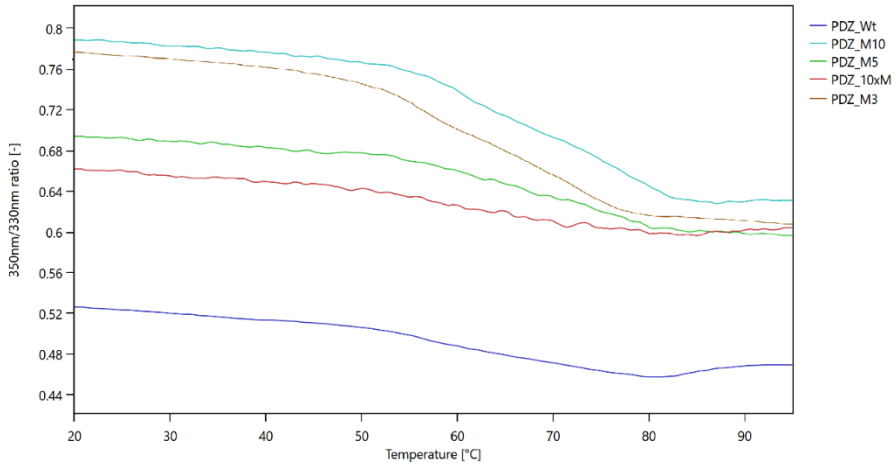


Figure 19 | **PDZ thermal unfolding.** Results of the qualitative measurements of wtPDZ, M10, 10xM, M5 and M3 using Prometheus NT.48. The concentration of the measured proteins was not standardized to the same value (in the order of µM). However, from the curve shapes I can state that the proteins are of sufficient quality for further experiments. The estimated denaturation midpoints were around 70-75 °C. (PDZ proteins do not contain Trp, however this does not prevent the use of Prometheus NT.48 (Misetic, 2016)).

I conducted the first binding experiment with wtPDZ-EGFP and max[FOXO4] = 28 µM and did not detected any binding event (Figure 20, brown line). We expected binding affinities to be in a range from µM to sub-mM. Maximal concentration of titrated molecule must be at least ten times higher than the actual binding affinity to be able to accurately determine K_d. Thus, I increased max concentration of FOXO4 ten times to 278 µM. Then I performed the binding experiment with wtPDZ again and detected a binding event (Figure 20, light blue line). I decided to use the same FOXO4 concentration in other binding experiments. However, even this concentration was not more than ten times higher than the actual binding affinities and therefore I was unable to determine an accurate K_d of PDZ-FOXO4 binding. I estimated the K_d values to be in a range from 100 µM to 10 mM and compared these binding affinities among predicted PDZ mutant variants and wtPDZ in relative terms. The highest binding affinity was observed for mutant variants carrying all predicted mutations and the lowest for wtPDZ (Figure 20). The negative control with EGFP protein was indeed negative. It is important to note that due to a shortage of funds I did not perform biological replicates for each PDZ domain, only one replication for wtPDZ, which was successful (Figure 21). Altogether I conducted seven binding experiments with PDZ domains, that were not influenced with the unknown fluorescence of non-tagged FOXO4 and all showed a positive signal compared to the negative control with EGFP.

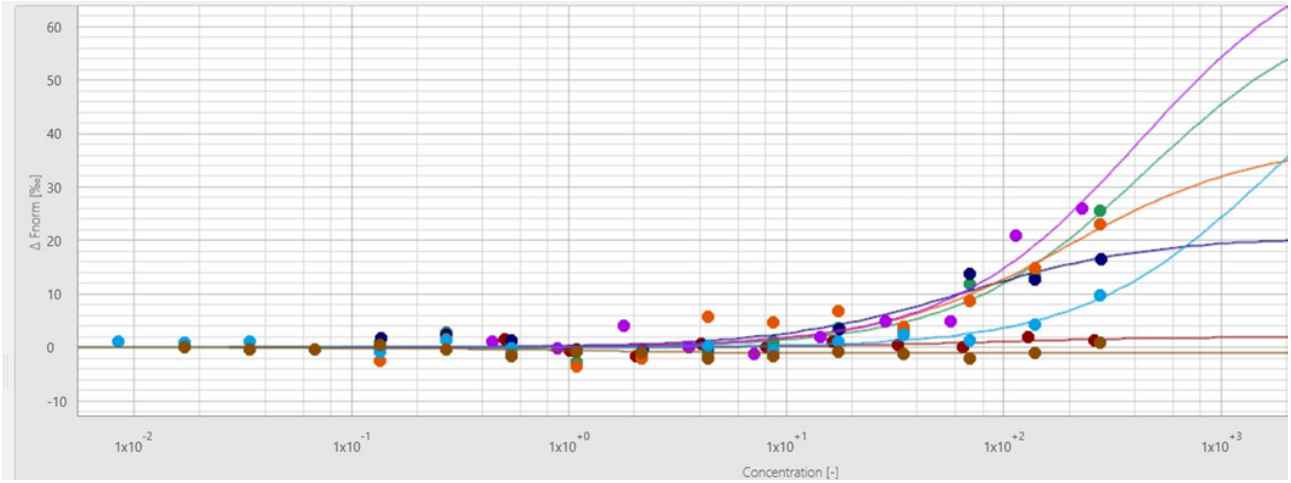


Figure 20 | **MST binding experiments.** *MST* binding experiments of predicted mutant variants PDZ – 10xM (green), M10 (orange), M5 (dark blue) and M3 (purple), wtPDZ (light blue) and wtPDZ with max[FOXO4] = 28 µM (brown). In all experiments, the highest concentration of FOXO4 was diluted with 2-fold serial dilution. Max [FOXO4] = *278 µM (*except 'brown' line). PDZ and EGFP were in nanomolar concentration.

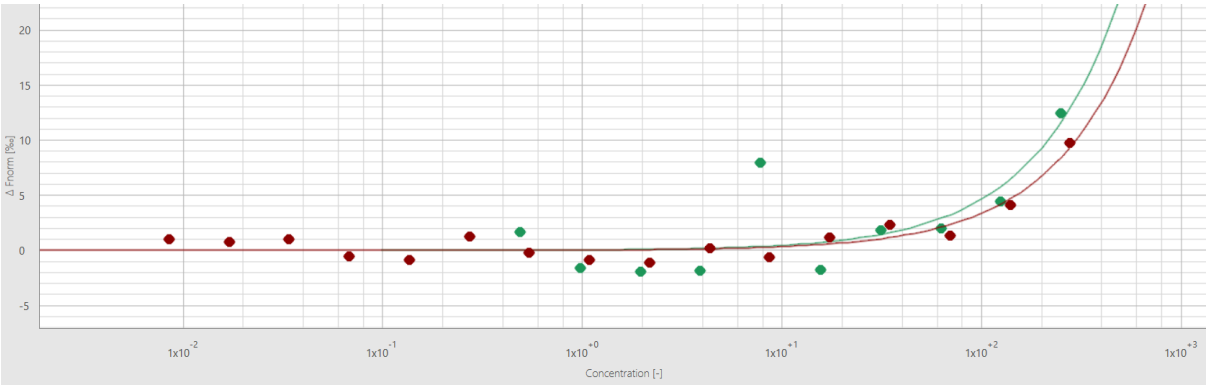


Figure 21 | **wtPDZ biological replicates.** MST binding experiments of two wtPDZ biological replicates. The highest concentration of FOXO4 [278 μM] was diluted with 2-fold serial dilution. wtPDZ were in nanomolar concentration.

4.2 Bioinformatic analysis of bladder cancer RNA-Seq data

In this part of my diploma thesis I analyzed differences between NP and P BCa with respect to cellular composition of TME and FOXO gene expression. In addition to planned tasks I decided to include an analysis of PI. Next, I conducted a pathway analysis and I investigated the relationship between cell types, PI and FOXO expression levels in individual BCa patients. This part of my thesis was performed at the Bioinformatics Research Center in Aarhus under supervision of Palle Villesen.

4.2.1 Bladder tumor microenvironment analysis from RNA-Seq data

I analyzed bladder tumor-infiltrating leukocytes from bulk tumor gene expression data using deconvolution method CIBERSORT. However, the quality of the deconvolution output and consequently information level for further analysis was poor (data not shown). I suggest it was caused by the fact that CIBERSORT was tested on microarray data (Newman et al., 2015). Fortunately, at the same time a new method called xCell for analyzing TME from bulk tumor gene expression data had emerged (Aran et al., 2017). This gene signature based method enables to infer 64 immune and stroma cell types providing more complex description of a TME. Additionally, xCell provides the information about the tumor purity defined as a microenvironment score.

4.2.1.1 Bladder tumor microenvironment analysis by xCell

Firstly, I compared the abundance of each cell type between P and NP BCa. Preliminary results of the xCell analysis showed only one significantly different ($p < .01$) cell type (Th1 cells) between NP and P BCa. Before the actual analysis I examined clinical information about the samples and I found out that the highest PFS in P group was 44 months (average = 11 m, median = 6 m) and in NP group 75 months (average = 36 m, median = 35 m). I performed the analysis with 12, 24 and 44 months PFS cutoff and found out that the number of significant results increased with higher PFS cutoff. I decided to use the highest cutoff (PFS > 44 months) for further analysis to reduce the noise from misclassified tumors to a minimum. Additionally, I excluded samples from patients treated with intravesical BCG vaccine and eventually performed xCell with 103 NP and 53 P NMIBC samples.

To test the difference in the cellular composition between NP and P I conducted a Welch Two Sample t-test. There was a significant difference ($p < .01$) in the amount of Th1 cells, erythrocytes, HSC, γδT cells, mesangial cells, CD4+ naive T-cells, Th2 cells, CLP and preadipocytes between NP and P BCa (Figure 22). There was no significant difference in the Microenvironment, Stroma and Immune score.

Since the presence of HSC in BCa is unlikely I decided to combine all stem cells and progenitor cells that are analyzed by xCell (HSC, CLP, CMP, GMP, MEP, MPP and MSC) as a marker for stemness in TME. I tested the difference in stemness marker between NP (sd = 0.092, mean = 0.066) and P (sd = 0.080, mean = 0.058) by Welch Two Sample t-test and the result was not significant ($p = 0.146$).

P_values	Cell_types	T_values	Mean_NP	Mean_P	Sd_NP	Sd_P
4.38E-07	Th1_cells	-5.516	3.21E-02	5.72E-02	2.04E-02	2.98E-02
3.05E-06	Erythrocytes	-5.081	2.52E-02	6.01E-02	2.59E-02	4.63E-02
5.03E-06	HSC	4.817	6.30E-02	3.90E-02	2.90E-02	2.97E-02
3.04E-05	Tgd_cells	-4.410	1.01E-02	1.57E-02	6.19E-03	8.11E-03
1.31E-04	Mesangial_cells	3.934	6.80E-02	4.18E-02	4.86E-02	3.37E-02
5.27E-04	CD4_naive_T_cells	3.554	1.57E-01	1.06E-01	9.94E-02	7.60E-02
4.57E-03	Th2_cells	2.899	3.88E-02	3.02E-02	1.75E-02	1.76E-02
5.03E-03	CLP	2.860	2.21E-01	1.70E-01	1.12E-01	1.01E-01
7.54E-03	Preadipocytes	2.721	8.49E-03	6.10E-03	5.47E-03	5.05E-03
2.86E-02	Neurons	2.218	1.17E-02	9.71E-03	5.50E-03	5.36E-03
3.91E-02	aDC	-2.104	1.91E-03	3.60E-03	2.96E-03	5.45E-03
4.17E-02	Eosinophils	2.054	7.57E-03	4.82E-03	1.01E-02	6.49E-03
4.81E-02	MPP	-2.012	5.66E-03	1.73E-02	2.15E-02	3.92E-02

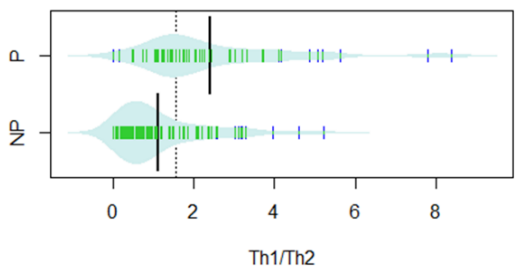


Figure 22 | **xCell result.** Table of differently abundant ($p < .05$) cell types between NP and P BCa (on the left) and beanplot of Th1/Th2 ratio in NP and P BCa (on the right).

4.2.1.2 Molecular classes and progression prediction from xCell data

To predict BCa progression and identify a cell type within TME with the highest predictive value I used three supervised machine learning algorithms – RF, SVM and NN. I firstly ran only RF and then SVM and NN were added to compare the performance of different machine learning algorithms on the given dataset. Since patient age, tumor grade and stage are known to influence disease outcomes I included them into the prediction models but without considerable improvement of prediction accuracy (Figure 23). Prediction models that included information about tumor grade contained less samples ('123 samples' group) due to a different grading system between dataset 1 (2004 WHO) and dataset 2 (1974 WHO). I took only grade 1 and grade 3 samples from dataset 2 and denoted them as low grade and high grade respectively. Prediction performance of these

A

		Validation						Test					
	Model	Confusion matrix		MCC	Accuracy	Sensitivity	Specificity	Confusion matrix		MCC	Accuracy	Sensitivity	Specificity
156 samples	Cells		NP	P	0.565016	0.809524	0.831461	0.756757		NP	P	0.533002	0.8	0.8182	0.75
		NP	74	9					NP	18	2				
		P	15	28					P	4	6				
	Model	Confusion matrix		MCC	Accuracy	Sensitivity	Specificity	Confusion matrix		MCC	Accuracy	Sensitivity	Specificity
	Cells+Age		NP	P	0.580624	0.81746	0.826087	0.794118		NP	P	0.533002	0.8	0.8182	0.75
		NP	76	7					NP	18	2				
		P	16	27					P	4	6				
123 samples	Model	Confusion matrix		MCC	Accuracy	Sensitivity	Specificity	Confusion matrix		MCC	Accuracy	Sensitivity	Specificity
	Cells+Age +Grade		NP	P	0.703418	0.868687	0.861111	0.888889		NP	P	0.408248	0.75	0.7778	0.6667
		NP	62	3					NP	14	2				
		P	10	24					P	4	4				
	Model	Confusion matrix		MCC	Accuracy	Sensitivity	Specificity	Confusion matrix		MCC	Accuracy	Sensitivity	Specificity
	Cells +Grade		NP	P	0.703418	0.868687	0.861111	0.888889		NP	P	0.408248	0.75	0.7778	0.6667
		NP	62	3					NP	14	2				
		P	10	24					P	4	4				
	Model	Confusion matrix		MCC	Accuracy	Sensitivity	Specificity	Confusion matrix		MCC	Accuracy	Sensitivity	Specificity
	Cells		NP	P	0.703418	0.868687	0.861111	0.888889		NP	P	0.4375	0.75	0.8125	0.625
		NP	62	3					NP	13	3				
		P	10	24					P	3	5				

B

		MDA			MDA
1	Th1_cells	0.02862	12	Astrocytes	0.00384
2	HSC	0.02347	13	Osteoblast	0.00383
3	Erythrocytes	0.01876	14	Macrophages_M2	0.00336
4	Tgd_cells	0.01738	15	CLP	0.00322
5	Age	0.00813	16	CD4_naive_T_cells	0.00275
6	CD4_memory_T_cells	0.00743	17	CD8_naive_T_cells	0.00263
7	Mast_cells	0.00617	18	Pericytes	0.00235
8	Chondrocytes	0.00615	19	Platelets	0.00165
9	Neurons	0.00568	20	Megakaryocytes	0.0013
10	Macrophages_M1	0.00542	21	aDC	0.00066
11	Skeletal_muscle	0.0041	22	Th2_cells	0.00053

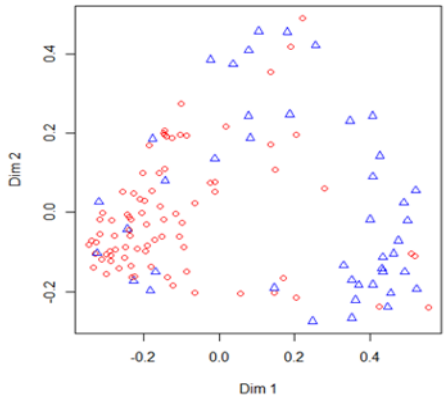


Figure 23 | **Random Forest results**. A | Results of five prediction models with the best performance in each category. Left: Estimated model performance. Right: Model performance on the test set. B | Left: Variable importance of the best 'Cells+age' prediction model. **Right:** MDS plot of the best 'Cells+Age' prediction model output. (Red circles – NP samples, blue triangles – P samples, MCC = Matthews Correlation Coefficient)

three machine learning algorithms were almost identical. The best RF prediction model for “156 samples” group achieved prediction accuracy 0.80 (MCC = 0.533) on the test set (Figure 23.A). Cell types used in this model are shown in Figure 23.B. They are ordered in a decreasing order of prediction accuracy measured with Mean Decrease Accuracy (MDA). In addition to cell types, patient age was used in this prediction model and was ranked as the fifth most important variable for the progression prediction. The highest prediction accuracy for RF prediction model for “123 samples” group was 0.75 (MCC = 0.438) (Figue 23.A).

The best SVM prediction model for the “156 samples” group achieved prediction accuracy 0.73 (MCC = 0.400) on the test set and 0.79 (MCC = 0.508) for the “123 samples” group. (see results in Supplementary information)

The best NN prediction model for the “156 samples” group predicted progression on the test set with 0.80 accuracy and 0.550 MCC, which was the best overall. For the “123 samples” group the prediction accuracy for the best model was 0.79 and MCC was 0.519. (see results in Supplementary information)

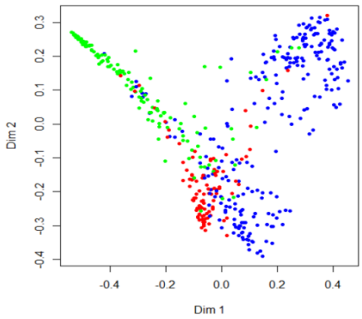
Then I exploited the cellular composition of TME for prediction of BCa molecular classes that were identified with 117-gene classifier (Figure 24.A). Even though the nature

A

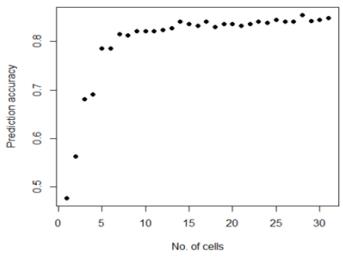
28 cells							
	C1	C2	C3		Sensitivity	Specificity	Accuracy
C1	73	17	7	Class 1	0.8588	0.9386	0.855
C2	9	228	12	Class 2	0.8571	0.9	
C3	3	21	106	Class 3	0.848	0.9316	

14 cells							
	C1	C2	C3		Sensitivity	Specificity	Accuracy
C1	69	21	7	Class 1	0.8313	0.9288	0.8403
C2	10	226	13	Class 2	0.8433	0.8894	
C3	4	21	105	Class 3	0.84	0.9288	

B



C



D

		MDA			MDA			MDA
1	Erythrocytes	0.04243	11	Osteoblast	0.01423	21	Neurons	0.00518
2	Tgd_cells	0.03351	12	Smooth_muscle	0.01208	22	Mast_cells	0.0047
3	CD4_naive_T_cells	0.03004	13	cDC	0.01101	23	CD4_T_cells	0.00444
4	Th1_cells	0.02857	14	CD4_memory_T_cells	0.00878	24	CD8_T_cells	0.00424
5	Epithelial_cells	0.02485	15	Chondrocytes	0.00878	25	Th2_cells	0.00369
6	Mesangial_cells	0.02056	16	Monocytes	0.00861	26	Macrophages_M2	0.00362
7	iDC	0.01993	17	Macrophages_M1	0.00858	27	Memory_B_cells	0.00358
8	Adipocytes	0.01897	18	Basophils	0.00722	28	Pericytes	0.00351
9	Astrocytes	0.01718	19	B_cells	0.00566	29	Macrophages	0.00329
10	Eosinophils	0.01538	20	NKT	0.00532	30	Fibroblasts	0.00324

Figure 24 **| Classes prediction with Random Forests. A |** RF results of two prediction models trained on data with 28 and 14 cell type variables. **B |** MDS plot of the ’28 cell types’ prediction model (class 1 – red, class 2 – blue, class 3 – green). **C |** Accuracy plot of prediction models (number of cell type variables used in model on x axis and prediction accuracy on y axis. **D |** Variable importance table showing the importance of each cell type for classes prediction.

of RF makes it robust against overfitting it should be taken into account that I did not split the dataset into validation- and test-set and thus the prediction results should be taken with a grain of salt. However, my main intention was to identify the cell types with the highest predictive value rather than predict the molecular classes (Figure 24).

4.2.2 Identification of para-inflammation in BCa from RNA-Seq data

Here I calculated PI score on the basis of signature of 40 genes and compared its value between NP and P BCa.

To assess the difference in the PI score between NP and P BCa patients I conducted a Welch Two Sample t-test. There was a significant difference ($p < .05$) in the PI score ($t = 6.167$, $p = 1.163e-08$) between NP ($m = 1.061$, $sd = 0.166$) and P BCa ($m = 0.895$, $sd = 0.156$) (Figure 25).

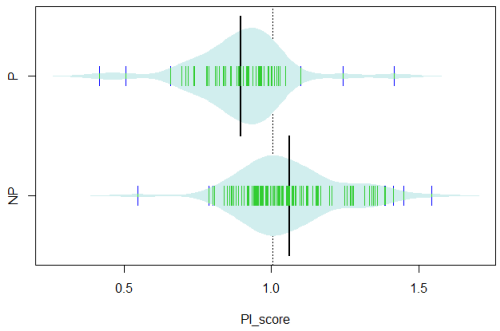


Figure 25 | **Beanplot of Pi scores in NP and P BCa patient cohors**.

4.2.3 FOXO focused expression analysis

The purpose of this analysis was to determine whether there was a difference between FOXO expression in normal bladder tissue and bladder cancer, NMIBC and MIBC, and NP and P BCa.

4.2.3.1 Differential expression between NMIBC, MIBC and normal tissue

To assess the difference between cancerous and normal tissue I exploited the expression data from the Genotype-Tissue Expression (GTEx) project (GTEx Consortium, 2013) (13 biological replicates) and defined significant results as a twofold expression level over/under the expression average. Expression level of FOXO3, FOXO4 and FOXO6 in NMIBC was 3.10, 2.64 and 2.05 lower respectively and expression level of FOXO1, FOXO3 and FOXO4 in MIBC was 2.53, 2.78 and 2.89 lower respectively compared to normal bladder tissue. To compare FOXO expression levels between NMIBC and MIBC I conducted a Welch Two Sample t-test. There was a statistically significant ($p < .05$) difference in the expression level of FOXO1 ($t=2.552$, $p=0.018$) between NMIBC ($m=5.700$, $sd=3.242$) and MIBC ($m=3.960$, $sd=3.199$) (Figure 26).

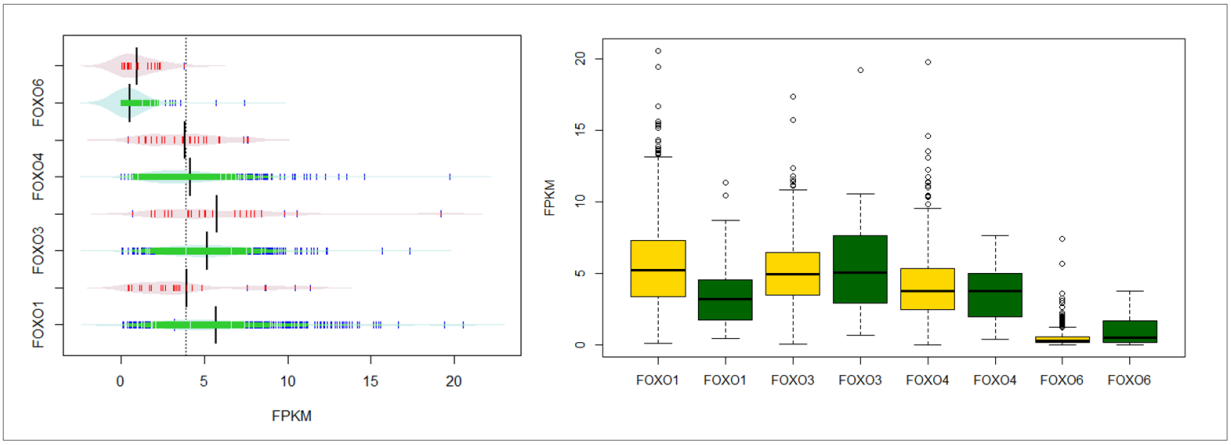


Figure 26 | **FOXO expression level in NMIBC and MIBC.** Beanplot (on the left) of FOXO expression level in NMIBC (green) and MIBC (red), and boxplot (on the right) of FOXO expression level in NMIBC (yellow) and MIBC (dark green).

In FOXO expression data in NMIBC many outliers can be observed (Figure 26). I explored the abnormal presence of high/low grade, high/low age and NP/P tumors in the outliers. I found a significantly abnormal ratio of high/low grade in the FOXO6 group (39 high/6 low grade vs 190 high/320 low grade) and in the FOXO4 group (8 high/6 low grade vs 190 high/320 low grade). This finding prompted me to compare FOXO expression between high grade and low grade in NMIBC samples with a Welch Two sample t-test. Indeed, there was a statistically significant difference ($p<.05$) in all FOXO expression levels – table on the right (HG – high grade, LG – low grade).

	P_values	T_values	Mean_HG	Mean_LG	Sd_HG	Sd_LG
FOXO1	4.30E-03	-2.870	5.133	5.993	3.353	3.136
FOXO3	4.07E-02	2.054	5.441	4.989	2.459	2.300
FOXO4	1.70E-02	2.399	4.486	3.938	2.624	2.251
FOXO6	3.01E-09	6.088	-1.276	-2.177	1.704	1.451

4.4.3.2 Differential expression between NP and P

Before the test I log transformed FOXO6 expression data and removed one outlier from FOXO6 in the NP group. Then I tested the difference in FOXO expression level between NP (PFS > 44 m, no BCG treatment) and P (no BCG treatment) using a Welch Two Sample t-test. There was a significant difference ($p < .05$) in FOXO1 ($t = 2.110$, $p = 0.037$) between NP ($m = 5.736$, $sd = 2.913$) and P ($m = 4.678$, $sd = 2.990$) and in FOXO6 ($t = -4.019$, $p = 1.166e-04$) between NP ($m = -2.461$, $sd = 1.573$) and P ($m = -1.310$, $sd = 1.753$) (Figure 27).

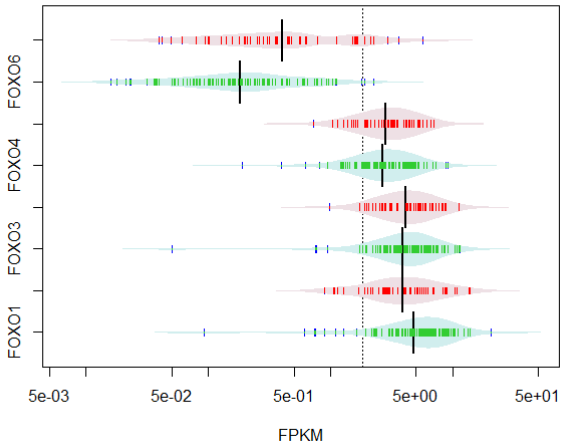


Figure 27 | **FOXO expression level in NP and P.** Beanplot of log transformed FOXO expression levels in NP (green) and P (red) BCa

4.4.4 Pathway analysis

Here, I performed an unbiased GSEA to investigate the cellular processes that differ between NP and P BCa. I was mainly interested in the analysis of metabolic pathways and in the relationship between FOXO expression levels and these pathways. This whole genome enrichment analysis showed enriched pathways involved in protein translation within NP group and in DNA replication, mitosis and cell cycle within P group. Nonetheless no metabolic pathway was shown to be enriched in NP or P BCa (Figure 28).

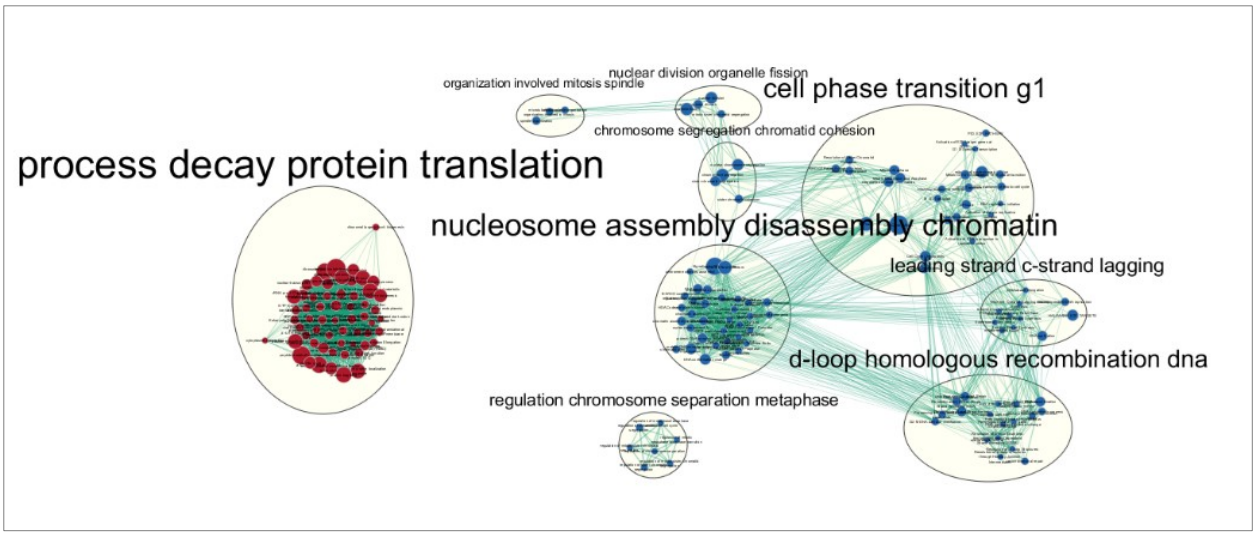


Figure 28 | **Visualization of enriched pathways in NP (red) and P (blue) BCa.**

4.4.5 Relationships between FOXO expression, TME and PI score in non-progressive and progressive BCa

Here I analyzed the relationship between FOXO expression levels, cellular compositions of TME and PI scores. To assess these correlations, I used the simplest approach – a Spearman's correlation test. I am aware of the naivety of this approach of assessing a correlation between one gene (FOXO) and cell type or PI score that are defined by multiple gene signatures and I did not expect high correlation coefficients.

Indeed, I did not find any statistically significant correlation ($p < .01$ and visual inspection) between FOXO expression and cell types in TME. Then I assessed the correlation between FOXO expression levels and PI scores. FOXO3 was statistically significantly positively correlated (rho = 0.225, p = 4.8 e-03) with PI score in NP+P dataset, but there was no statistically significant correlation in NP or P group only. Next, I explored the correlation between PI scores and cell types that differ between NP and P. In NP+P Th1 cell type was statistically significantly negatively correlated with PI score (rho = -0.213, p = 7.5 e-03).

5 Discussion

In the first experimental part of my thesis, using MST binding assays I proved that the PDZ domain structure is suitable for rational design. During the binding experiment, I detected a fluorescence signal from the non-tagged FOXO4-DBD protein. Firstly, on the basis of spectrofluorimetry results I thought that I accidentally contaminated the FOXO4 sample with PDZ-EGFP fusion protein. Nevertheless, after purification of the FOXO sample by IEC I still detected the fluorescence. I observed the highest fluorescence in the first eluted fraction from IEC column which gradually decreased with increasing ion concentration independently on protein concentration. I hypothesized that the fluorescence could be caused by the PDZ-EGFP protein bound to FOXO4. But then, this complex would be removed by SEC and additionally would be detected by SDS-PAGE and/or MALDI-TOF analysis. Since I identified degradation products of FOXO4 which have a slightly different net surface charge it could explain why I detected the highest fluorescence in the first eluted fraction from IEC. I hypothesize that the fluorescence is the intrinsic property of some degradation product of FOXO4. However, taken together, the source of fluorescence remains elusive.

Rational design is one approach of generating high-affinity binding proteins for a given target. Key interface residues and consequently beneficial mutations are identified on the basis of structural and/or functional knowledge of protein molecules. The advantage of this computational method is the reduction of an extremely large number of protein sequences to a few which can be experimentally tested. The major limitation of rational design is its relatively low prediction accuracy of the mutations' effect on protein function, which is caused by complex relationships between the amino acid sequence and protein secondary structure. In my thesis, I verified the bioinformatic prediction of PDZ binding to FOXO4-DBD domain and the choice of mutant variants. However, I showed that the estimated binding affinities are very low (in the range from 100 μM to 10 mM). Affinity of FOXO4-DBD to dsDNA is 188 ± 10 nM (Boura et al., 2007), therefore, to be able to efficiently inhibit this binding we have to increase the affinity of our PDZ domain at least a thousand times.

An alternative way of engineering binding proteins is a combinatorial technique which mimics affinity maturation of antibodies (also known as *in vitro* directed evolution). In this method random mutations are introduced into DNA sequences and the proteins with favorable mutations are selected by display technologies, such as bacterial display (Löfblom, 2011). This process can be repeated until the protein with the required function

(e.g. high binding affinity for a given target) is selected. Both rational design and combinatorial approaches have their pros and cons, therefore, the best strategy for developing high-affinity binding proteins is to combine these two approaches in order to leverage their advantages (i.e. semi-rational approach). For example, Karanicolas et al. (2011) used the computational method to design a novel binding interface between two non-interacting proteins. Their best PPI design possessed a K_d of ~130 nM, which was then increased by directed evolution using yeast surface display to a K_d of ~180 pM.

There are two possible future ways how to increase the PDZ domain binding affinity to FOXO4-DBD. One way is to analyze the PDZ/FOXO4 binding interface by two complementary experimental methods HDX-MS and NMR. Experimentally obtained data would then be used for the improvement of the prediction algorithm and new predicted mutant variants would be experimentally verified. The other way is to include a combinatorial protein engineering and then analyze the binding interface of high affinity variants to improve the algorithm for *in silico* PPI prediction based on experimentally obtained data. Given the evidence in literature, I think the latter approach is better and therefore I will perform it as a follow-up experiment.

In the second bioinformatic part of my thesis, I described the differences in TME, PI score and FOXO genes expression levels in NMIBC (diagnosed as the first tumor) that are associated with later disease progression to MIBC. Then I explored the relationship between FOXO expression, TME and PI score based on bulk tumor RNA-Seq expression data.

Altogether I collected 510 NMIBC RNA-Seq samples from which ~11 % (57) were samples from progressive tumors (T2+ progression). This corresponds to the expected 5-year progression rate of 10-15 % (Knowles and Hurst, 2015). The limitation of this study is that not all patients were followed for the period of five years (Hedegaard et al., 2016). The analysis of TME between 453 NP and 57 P BCa showed only one significantly different cell type (Th1 cell). The average PFS rate within these 453 NP samples was 36 months (median = 35 m, max = 75 months). I analyzed the differences in TME between P samples and NP samples with different PFS cutoffs (12, 24 and 44 months). I found that the number of significantly different cell types increased with higher PFS cutoff. Thus, in further analysis I used non-progressive samples with PFS > 44 months to reduce the noise from misclassified tumors to a minimum. This bias towards patients with high PFS within non-progressive group is another limitation of this study. Additionally, I excluded samples

from patients treated with intravesical BCG vaccine, because not all patients suitable for BCG treatment received it and therefore it could influence the outcomes for these patients.

The most significantly different cell types between NP (PFS > 44 months) and P in TME characterized in this thesis were Th1 cells, erythrocytes and γδT cells. Their negative prognostic value in BCa was identified by t-test and validated with a Random Forest. In addition, HSC appeared to significantly differ between NP and P in TME. The presence of HSC in bladder cancer has not been described in literature and is unlikely. I rather suggest the use of all stem cells and progenitor cells that are analyzed by xCell as a marker for stemness in the TME and I found that the difference in this marker between NP and P BCa was not significant.

Another cell type with a negative prognostic value were erythrocytes. Their elevated level could be explained by increased levels of angiogenesis which is one of the hallmarks of cancer (Hanahan and Weinberg, 2011) or by increased tumor blood vessel leakiness. Nevertheless, their higher leakiness would lead to higher infiltration of TME with immune cells, but xCell analysis did not show the difference in the abundance of tumor infiltrating leukocytes (defined as Immune score in xCell) between NP and P BCa. This suggests that in this case the higher level of erythrocytes in P BCa is caused by increased angiogenesis forming a denser tumor vasculature.

Previous studies have suggested that both Th1 and Th2 cell types mediate antitumor immunity, although it was shown that Th2 antitumor immunity is less potent and may be context dependent. The IL-4 cytokine secreted by Th2 cells probably exerts antitumor effects via enhanced infiltration of innate cells, including eosinophils (reviewed in (Kim and Cantor, 2014)). Indeed, I observed higher levels of eosinophils in NP compared to P but with a less-stringent p-value cutoff ($p<0.05$). Their natural low abundance makes it difficult to detect them with xCell method which led to many zero values. These constant zero values could be the reason why Random Forest algorithm did not select them as important cells for the prediction model. To my knowledge, there is no article describing pro-tumorigenic effect of Th1 cells which I found to be more abundant in P compare to NP BCa in this thesis. Interestingly, antitumor cells (CD8 cells and macrophages) that are regulated by Th1 cells did not differ in TME between NP and P BCa. I suggest that in this context the Th1/Th2 ratio is a more important factor for BCa progression than just Th1 or Th2 cell abundance alone. Recent findings described the effect of CD4+ T cells on vessel normalization (Tian et al., 2017). This implies that the ratio of Th1/Th2 could also play important role in angiogenesis.

For a long time, γδT cells were recognized as cells with anti-tumor properties. Additionally, the authors of CIBERSORT have showed that γδT cells were the best positive predictive parameters across different cancer types, BCa included (Gentles et al., 2015). But several recent studies indicate pro-tumorigenic effect of γδT cells by various mechanism, for example by IL-17 production which promotes tumor angiogenesis (Wakita et al., 2010). In this thesis, I described their negative prognostic value in BCa using xCell analysis.

I was able to construct the NN prediction model for NMIBC progression with an 80 % prediction accuracy using the information about patient age and cellular composition of TME from xCell analysis. The EORTC (European Organization for Research and Treatment of Cancer) risk tables (Sylvester et al., 2006) and the CUETO (Spanish Urological Club for Oncological Treatment) scoring model (Fernandez-Gomez et al., 2009) are the best-established prediction tools for prediction of NMIBC recurrence and progression. The EORTC risk table is based on tumor stage and grade, number of tumors, tumor size, concomitant CIS, and history of prior disease recurrence. Its prediction accuracy of 1-year and 5-year PFS was 74 % and 75 % respectively (validated both internally and externally). The CUETO scoring model is based on age, gender, prior recurrence status, number of tumors, tumor stage, tumor grade, and the presence of concomitant CIS. Its prediction accuracy of 1-year and 5-year PFS was 69 % and 70 % respectively (validated both internally and externally) (Kluth et al., 2015). I think that my prediction model would not outperform the EORTC risk tables or the CUETO in the clinic. Nevertheless, results presented in my thesis add further evidence to the importance of TME in BCa progression and to the fact that microenvironment signatures can be exploit by prognostic and prediction tools in BCa management. In my opinion, to be able to stratify patient treatment and their monitoring accurately, a combination of multiple markers and even inclusion of those that are not yet known to influence BCa progression will have to be applied. For example, the unpublished data (Bao et al., 2016) from the laboratory of Jeremy Burton (London, Canada) imply the importance of the bladder microbiome in BCa recurrence.

In addition, I showed that the cellular composition of TME could predict molecular classes of BCa with an 85 % prediction accuracy. Recently, Becht et al. (2016a) demonstrated that microenvironment signatures correlate with molecular subtypes of colorectal cancer. They analyzed the TME by Microenvironment Cell Populations-counter computational algorithm, which is able to infer the absolute abundance of eight immune

(CD3+ T cells, CD8+ T cells, cytotoxic lymphocytes, NK cells, B lymphocytes, cells originating from monocytes, myeloid dendritic cells, neutrophils) and two stromal cell types (endothelial cells, fibroblasts) from transcriptomic data (Becht et al., 2016b). Subsequently, they validated their results using immunohistochemistry (Becht et al., 2016a). The validation and description of which cell types are typical for certain BCa molecular classes was beyond the scope of my project but it will be later followed in future studies in my home laboratory.

While xCell outperforms other methods for *in silico* analysis of TME, its correlations are still far from perfect with direct measurements (Aran et al., 2017) and this could be the cause of the relatively high error rates of my prediction models. But on the other hand, it might also cause false positive results and thus findings presented in this thesis must be experimentally validated using single cell sequencing and/or single cell proteomic analysis.

The higher PI score observed in NP BCa in comparison to P BCa is in concert with the results of the paper defining PI score (Aran et al., 2016). The authors linked elevated levels of PI score to increased mortality in most cancers with the exception of chromophobe renal cell carcinoma, prostate cancer and bladder cancer. Interestingly, these are the tumors of urinary tract, which indicates that this environment may play the important role in PI. Even though it was previously shown that nonsteroidal anti-inflammatory drugs could prevent BCa (Daugherty et al., 2011) and reduce the risk of BCa recurrence (Gee et al., 2009), results presented in this thesis and in (Aran et al., 2016) suggest that low-grade inflammation, coined as para-inflammation, might be a positive prognostic marker in BCa.

Cheah et al. (2015) showed that bladder tumors bearing CD14-high cancer cells are bigger and form a denser tumor vasculature with higher myeloid cell infiltration. In my thesis, I showed a higher abundance of erythrocytes in P BCa compared to NP BCa, which corresponds to this finding. These CD14-high cancer cells express a high level of numerous inflammatory mediators and therefore might play a key role in tumor-promoting inflammation. However, CD14 is also a part of PI gene signature. In addition, the expression level of CD14 did not significantly differ between NP and P BCa in my dataset.

The expression levels of FOXO transcription factors were lower in BCa compared to normal bladder tissue expression levels from GTEx. This was expected since in most cases they are considered to function as tumor suppressors (Brachène and Demoulin,

2016). FOXO4, which was chosen as the target for inhibition described in this thesis, was found to have lower expression levels in low grade BCa but differences in its expression were not associated with BCa progression. Recently FOXO1 down-regulation was associated with worse outcome in BCa (Lloreta et al., 2017). Indeed, I found lower expression levels of FOXO1 to be associated with high tumor grade and progression. In addition, FOXO1 was shown to suppress gastric cancer growth via inhibition of angiogenesis (Kim et al., 2016). This corresponds to findings presented in this thesis and could possibly explain the positive prognostic effect of FOXO1 in BCa. Also, FOXO6 overexpression was recently linked to progression and prognosis in gastric cancer (Wang et al., 2017). Here I showed its higher levels to be associated with P BCa.

In this thesis, I described the prognostic value of the cellular composition of TME in BCa. Based on my results and on given literature evidence I assume that angiogenesis (influenced by Th1/Th2 ratio, γδT cells and FOXO1 expression levels) has an important role in later NMIBC progression to MIBC. I hypothesize that Th1/Th2 ratio also affects PI. This is supported by my finding of a weak negative correlation between Th1 and PI. In context of BCa PI has a positive prognostic value. I think that the environment of the urinary tract has a significant effect on PI and distinguishes BCa from other cancer types.

6 Supplementary information

6.1 Primer sequences for mutagenesis

P_V20Q/1

CGT CGT GTG GAT GAG CCG GGT GTG Tm = 64 C

V20Q/2

TTG CAG TTT GAT ACC CAG TTG CTC C Tm = 64 C

P_R22D/1

GTG GAT GAG CCG GGT G Tm = 64 C

R22D/2

ATC ACG CAC CAG TTT GAT ACC C Tm = 63 C

P_V23T/1

GAT GAG CCG GGT GTG TTC Tm = 64 C

V23T/2

AGT ACG ACG CAC CAG TTT GAT AC Tm = 63 C

P_F31Y/1

AAC GTG CTG AAC GGT GG Tm = 64 C

F31Y/2

ATA GAT GAA CAC ACC CGG CTC Tm = 64 C

P_N32E/1

GTG CTG AAC GGT GGC Tm = 62 C

N32E/2

TTC GAA GAT GAA CAC ACC CGG C Tm = 64 C

P_L34F/1

AAC GGT GGC GTT GCG Tm = 65 C

L34F/2

AAA CAC GTT GAA GAT GAA CAC ACC C Tm = 64 C

P_N35E/1

GGT GGC GTT GCG GAC Tm = 65 C

N35E/2

TTC CAG CAC GTT GAA GAT GAA CAC A Tm = 64 C

P_E46K/1

GAA AAC GAT CGT GTG CTG GC Tm = 65 C

E46K/2

TTT CAG CTG ACC GTG ACG G Tm = 64 C

P_E47I/1

AAC GAT CGT GTG CTG GC Tm = 64 C

E47I/2

AAT CTC CAG CTG ACC GTG AC Tm = 63 C

P_N48F/1

GAT CGT GTG CTG GCG ATT AA Tm = 63 C

N48F/2

AAA TTC CTC CAG CTG ACC GT Tm = 64 C

6.2 wtPDZ-EGFP and FOXO4-DBD protein sequences

wtPDZ

```
          10         20         30         40         50         60
MSYYHHHHHH   DYDIPTTENL   YFQGAMNVSK   GEELFTGVVP   ILVELDGDVN   GHKFSVSGEG

          70         80         90        100        110        120
EGDATYGKLT   LKFICTTGKL   PVPWPTLVTT   LTYGVQCFSR   YPDHMKQHDF   FKSAMPEGYV

         130        140        150        160        170        180
QERTIFFKDD   GNYKTRAEVK   FEGDTLVNRI   ELKGIDFKED   GNILGHKLEY   NYNSHNVYIM

         190        200        210        220        230        240
ADKQKNGIKV   NFKIRHNIED   GSVQLADHYQ   QNTPIGDGPV   LLPDNHYLST   QSALSKDPNE

         250        260        270        280        290        300
KRDHMVLLEF   VTAAGITLGM   DELYKAMGSF   HVILNKSSPE   EQLGIKLVRR   VDEPGVFIFN

         310        320        330        340        350
VLNGGVADRH GQLEENDRVL AINGHDLRFG SPESAAHLIQ ASERRVHLVV SRQ
```

Yellow – wtPDZ
Green – EGFP
Cyan – His-tag
Gray – TEV-site

Number of amino acids: 353

Molecular weight: 39869.8

Theoretical pI: 5.90

FOXO4-DBD

```
         10         20         30         40         50         60
GSSHHHHHHS SGLVPRGSHM LEDPGAVTGP RKGGSRRNAW GNQSYAELIS QAIESAPEKR

         70         80         90        100        110        120
LTLAQIYEWM VRTVPYFKDK GDSNSSAGWK NSIRHNLSLH SKFIKVHNEA TGKSSWWMLN

        130        140        150
PEGGKSGKAP RRRAASMDSS SKLLRGRSKA
```

Yellow - 82G - FOXO4 - 207A
Cyan - His-tag

Number of amino acids: 150

Molecular weight: 16577.58

Theoretical pI: 10.53

6.3 Para-inflammation gene signature

(Copied from Additional file 3: Table S1 of (Aran et al., 2016))

Mouse gene symbol	Human gene symbol	Entrez ID	Entrez Gene Name	Type(s)
Aim2	AIM2	9447	absent in melanoma 2	other
Anxa1	ANXA1	301	annexin A1	enzyme
Blnk	BLNK	29760	B-cell linker	other
Bst2	BST2	684	bone marrow stromal cell antigen 2	other
Ccnd1	CCND1	595	cyclin D1	transcription regulator
Cd14	CD14	929	CD14 molecule	transmembrane receptor
Cd276	CD276	80381	CD276 molecule	other
Cd44	CD44	960	CD44 molecule (Indian blood group)	enzyme
Cxcl10	CXCL10	3627	chemokine (C-X-C motif) ligand 10	cytokine
Cxcl9	CXCL9	4283	chemokine (C-X-C motif) ligand 9	cytokine
Hmox1	HMOX1	3162	heme oxygenase 1	enzyme
Icam1	ICAM1	3383	intercellular adhesion molecule 1	transmembrane receptor
Ifit1	IFIT1	3434	interferon-induced protein with tetratricopeptide repeats 1	other
Ifit2	IFIT2	3433	interferon-induced protein with tetratricopeptide repeats 2	other
Ifit3	IFIT3	3437	interferon-induced protein with tetratricopeptide repeats 3	other
Ifitm3	IFITM3	10410	interferon induced transmembrane protein 3	other
Il1rn	IL1RN	3557	interleukin 1 receptor antagonist	cytokine
Il33	IL33	90865	interleukin 33	cytokine
Isg15	ISG15	9636	ISG15 ubiquitin-like modifier	other
Itga2	ITGA2	3673	integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)	transmembrane receptor
Lgmn	LGMN	5641	legumain	peptidase

Mmp7	MMP7	4316	matrix metallopeptidase 7	peptidase
Mx1	MX1	4599	MX dynamin-like GTPase 1	enzyme
Mx2	MX2	4600	MX dynamin-like GTPase 2	enzyme
Nox1	NOX1	27035	NADPH oxidase 1	ion channel
Oas1g	OAS1	4938	2'-5'-oligoadenylate synthetase 1, 40/46kDa	enzyme
Oas2	OAS2	4939	2'-5'-oligoadenylate synthetase 2, 69/71kDa	enzyme
Oas3	OAS3	4940	2'-5'-oligoadenylate synthetase 3, 100kDa	enzyme
Pla2g2a	PLA2G2A	5320	phospholipase A2, group IIA (platelets, synovial fluid)	enzyme
Pla2g2d	PLA2G2D	26279	phospholipase A2, group IID	enzyme
Plat	PLAT	5327	plasminogen activator, tissue	peptidase
Plaur	PLAUR	5329	plasminogen activator, urokinase receptor	transmembrane receptor
Pparg	PPARG	5468	peroxisome proliferator-activated receptor gamma	ligand-dependent nuclear receptor
Ptges	PTGES	9536	prostaglandin E synthase	enzyme
Rel	REL	5966	v-rel avian reticuloendotheliosis viral oncogene homolog	transcription regulator
Retnlb	RETNLB	84666	resistin like beta	other
Scarb1	SCARB1	949	scavenger receptor class B, member 1	transporter
Tirap	TIRAP	114609	toll-interleukin 1 receptor (TIR) domain containing adaptor protein	other
Tlr2	TLR2	7097	toll-like receptor 2	transmembrane receptor
Tnfrsf12a	TNFRSF12A	51330	tumor necrosis factor receptor superfamily, member 12A	transmembrane receptor

6.4 SVM and NN results

A

		10x Cross-validation							Test							
	Model	**Confusion matrix**			**MCC**	**Accuracy**	**Sensitivity**	**Specificity**	**Confusion matrix**			**MCC**	**Accuracy**	**Sensitivity**	**Specificity**
			NP	P						NP	P				
156 samples	Cells	NP	82	1	0.85905	0.93651	0.92135	0.97297	NP	16	4	0.4	0.73333	0.8	0.6
		P	7	36					P	4	6				
	Model	**Confusion matrix**			**MCC**	**Accuracy**	**Sensitivity**	**Specificity**	**Confusion matrix**			**MCC**	**Accuracy**	**Sensitivity**	**Specificity**
	Cells+Age		NP	P						NP	P				
		NP	82	1	0.80689	0.9127	0.8913	0.97059	NP	16	4	0.30861	0.7	0.7619	0.55556
		P	10	33					P	5	5				
123 samples	**Model**	**Confusion matrix**			**MCC**	**Accuracy**	**Sensitivity**	**Specificity**	**Confusion matrix**			**MCC**	**Accuracy**	**Sensitivity**	**Specificity**
	Cells+Age+Grade		NP	P						NP	P				
		NP	64	1	0.86542	0.93939	0.92754	0.96667	NP	14	2	0.40825	0.75	0.77778	0.66667
		P	5	29					P	4	4				
	Model	**Confusion matrix**			**MCC**	**Accuracy**	**Sensitivity**	**Specificity**	**Confusion matrix**			**MCC**	**Accuracy**	**Sensitivity**	**Specificity**
	Cells +Grade		NP	P						NP	P				
		NP	64	1	0.7547	0.88889	0.86486	0.96	NP	15	1	0.50783	0.79167	0.78947	0.8
		P	10	24					P	4	4				
	Model	**Confusion matrix**			**MCC**	**Accuracy**	**Sensitivity**	**Specificity**	**Confusion matrix**			**MCC**	**Accuracy**	**Sensitivity**	**Specificity**
	Cells		NP	P						NP	P				
		NP	63	2	0.88744	0.94949	0.95455	0.93939	NP	13	3	0.20412	0.66667	0.72222	0.5
		P	3	31					P	5	3				

B

		Validation							Test							
	Model	**Confusion matrix**			**MCC**	**Accuracy**	**Sensitivity**	**Specificity**	**Confusion matrix**			**MCC**	**Accuracy**	**Sensitivity**	**Specificity**
			NP	P						NP	P				
156 samples	Cells	NP	21	0	0.79772	0.90625	0.875	1	NP	17	3	0.17678	0.66667	0.70833	0.5
		P	3	8					P	7	3				
	Model	**Confusion matrix**			**MCC**	**Accuracy**	**Sensitivity**	**Specificity**	**Confusion matrix**			**MCC**	**Accuracy**	**Sensitivity**	**Specificity**
	Cells+Age		NP	P						NP	P				
		NP	20	1	0.86147	0.9375	0.95238	0.90909	NP	17	3	0.55	0.8	0.85	0.7
		P	1	10					P	3	7				
123 samples	**Model**	**Confusion matrix**			**MCC**	**Accuracy**	**Sensitivity**	**Specificity**	**Confusion matrix**			**MCC**	**Accuracy**	**Sensitivity**	**Specificity**
	Cells+Age+Grade		NP	P						NP	P				
		NP	16	0	0.90749	0.95833	0.94118	1	NP	13	3	0.4375	0.75	0.8125	0.625
		P	1	7					P	3	5				
	Model	**Confusion matrix**			**MCC**	**Accuracy**	**Sensitivity**	**Specificity**	**Confusion matrix**			**MCC**	**Accuracy**	**Sensitivity**	**Specificity**
	Cells +Grade		NP	P						NP	P				
		NP	16	0	0.8165	0.91667	0.88889	1	NP	12	4	0.36515	0.70833	0.8	0.55556
		P	2	6					P	3	5				
	Model	**Confusion matrix**			**MCC**	**Accuracy**	**Sensitivity**	**Specificity**	**Confusion matrix**			**MCC**	**Accuracy**	**Sensitivity**	**Specificity**
	Cells		NP	P						NP	P				
		NP	16	0	0.72548	0.875	0.84211	1	NP	14	2	0.51856	0.79167	0.82353	0.71429
		P	3	5					P	3	5				

Figure: Support Vector Machine and Neural Network results A. Results of five SVM prediction models. **Left:** Estimated model performance. **Right:** Model performance on the test set. **B.** Results of five NN prediction models **Left:** Estimated model performance. **Right:** Model performance on the test set.

7 References

Aran, D., Hu, Z., Butte, A.J., 2017. xCell: Digitally portraying the tissue cellular heterogeneity landscape. bioRxiv. doi:10.1101/114165

Aran, D., Lasry, A., Zinger, A., Biton, M., Pikarsky, E., Hellman, A., Butte, A.J., Ben-Neriah, Y., 2016. Widespread parainflammation in human cancer. Genome Biol. 17, 145. doi:10.1186/s13059-016-0995-z

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A.F., Nielsen, H., 2000. Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics 16, 412–424. doi:10.1093/bioinformatics/16.5.412

Banta, S., Dooley, K., Shur, O., 2013. Replacing Antibodies: Engineering New Binding Proteins, in: Yarmush, M.L. (Ed.), Annual Review of Biomedical Engineering, Vol 15. Annual Reviews, Palo Alto, pp. 93–113.

Bao, Y., Razvi, H., Gloor, G., Reid, G., Denstedt, J., Burton, J., 2016. MP88-19 URINARY MICROBIOME PATTERNS APPEARS TO BE CORRELATIVE TO INTRAVESICAL RECURRENCE OF NON-MUSCLE INVASIVE BLADDER CANCER. J. Urol. 195, e1136. doi:10.1016/j.juro.2016.02.2437

Becht, E., de Reyniès, A., Giraldo, N.A., Pilati, C., Buttard, B., Lacroix, L., Selves, J., Sautès-Fridman, C., Laurent-Puig, P., Fridman, W.H., 2016a. Immune and Stromal Classification of Colorectal Cancer Is Associated with Molecular Subtypes and Relevant for Precision Immunotherapy. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 22, 4057–4066. doi:10.1158/1078-0432.CCR-15-2879

Becht, E., Giraldo, N.A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., Selves, J., Laurent-Puig, P., Sautès-Fridman, C., Fridman, W.H., de Reyniès, A., 2016b. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. 17, 218. doi:10.1186/s13059-016-1070-5

Becker, T., Loch, G., Beyer, M., Zinke, I., Aschenbrenner, A.C., Carrera, P., Inhester, T., Schultze, J.L., Hoch, M., 2010. FOXO-dependent regulation of innate immune homeostasis. Nature 463, 369–373. doi:10.1038/nature08698

Boura, E., Rezabkova, L., Brynda, J., Obsilova, V., Obsil, T., 2010. Structure of the human FOXO4-DBD-DNA complex at 1.9 Å resolution reveals new details of FOXO binding to the DNA. Acta Crystallogr. D Biol. Crystallogr. 66, 1351–1357. doi:10.1107/S0907444910042228

Boura, E., Silhan, J., Herman, P., Vecer, J., Sulc, M., Teisinger, J., Obsilova, V., Obsil, T., 2007. Both the N-terminal loop and wing W2 of the forkhead domain of transcription factor FoxO4 are important for DNA binding. J. Biol. Chem. 282, 8265–8275. doi:10.1074/jbc.M605682200

Brachène, A.C. de, Demoulin, J.-B., 2016. FOXO transcription factors in cancer development and therapy. Cell. Mol. Life Sci. 73, 1159–1172. doi:10.1007/s00018-015-2112-y

Brunet, A., Bonni, A., Zigmond, M.J., Lin, M.Z., Juo, P., Hu, L.S., Anderson, M.J., Arden, K.C., Blenis, J., Greenberg, M.E., 1999. Akt Promotes Cell Survival by Phosphorylating and Inhibiting a Forkhead Transcription Factor. Cell 96, 857–868. doi:10.1016/S0092-8674(00)80595-4

Brunet, A., Park, J., Tran, H., Hu, L.S., Hemmings, B.A., Greenberg, M.E., 2001. Protein kinase SGK mediates survival signals by phosphorylating the forkhead transcription factor FKHRL1 (FOXO3a). Mol. Cell. Biol. 21, 952–965. doi:10.1128/MCB.21.3.952-965.2001

Calnan, D.R., Brunet, A., 2008. The FoxO code. Oncogene 27, 2276–2288. doi:10.1038/onc.2008.21

Charitou, P., Rodriguez-Colman, M., Gerrits, J., van Triest, M., Groot Koerkamp, M., Hornsveld, M., Holstege, F., Verhoeven-Duif, N.M., Burgering, B.M.T., 2015. FOXOs support the metabolic requirements of normal and tumor cells by promoting IDH1 expression. EMBO Rep. 16, 456–466. doi:10.15252/embr.201439096

Cheah, M.T., Chen, J.Y., Sahoo, D., Contreras-Trujillo, H., Volkmer, A.K., Scheeren, F.A., Volkmer, J.-P., Weissman, I.L., 2015. CD14-expressing cancer cells establish the inflammatory and proliferative tumor microenvironment in bladder cancer. Proc. Natl. Acad. Sci. U. S. A. 112, 4725–4730. doi:10.1073/pnas.1424795112

Cheung, R.C.F., Wong, J.H., Ng, T.B., 2012. Immobilized metal ion affinity chromatography: a review on its applications. Appl. Microbiol. Biotechnol. 96, 1411–1420. doi:10.1007/s00253-012-4507-0

Choi, W., Porten, S., Kim, S., Willis, D., Plimack, E.R., Hoffman-Censits, J., Roth, B., Cheng, T., Tran, M., Lee, I.-L., Melquist, J., Bondaruk, J., Majewski, T., Zhang, S., Pretzsch, S., Baggerly, K., Siefker-Radtke, A., Czerniak, B., Dinney, C.P.N., McConkey, D.J., 2014. Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. Cancer Cell 25, 152–165. doi:10.1016/j.ccr.2014.01.009

Chung, S., Lee, T.J., Reader, B.F., Kim, J.Y., Lee, Y.G., Park, G.Y., Karpurapu, M., Ballinger, M.N., Qian, F., Rusu, L., Chung, H.Y., Unterman, T.G., Croce, C.M., Christman, J.W., 2016. FoxO1 regulates allergic asthmatic inflammation through regulating polarization of the macrophage inflammatory phenotype. Oncotarget 7, 17532–17546. doi:10.18632/oncotarget.8162

Daitoku, H., Sakamaki, J.-I., Fukamizu, A., 2011. Regulation of FoxO transcription factors by acetylation and protein-protein interactions. Biochim. Biophys. Acta 1813, 1954–1960. doi:10.1016/j.bbamcr.2011.03.001

Daugherty, S.E., Pfeiffer, R.M., Sigurdson, A.J., Hayes, R.B., Leitzmann, M., Schatzkin, A., Hollenbeck, A.R., Silverman, D.T., 2011. Nonsteroidal antiinflammatory drugs and bladder cancer: a pooled analysis. Am. J. Epidemiol. 173, 721–730. doi:10.1093/aje/kwq437

Dejean, A.S., Hedrick, S.M., Kerdiles, Y.M., 2011. Highly Specialized Role of Forkhead Box O Transcription Factors in the Immune System. Antioxid. Redox Signal. 14, 663–674. doi:10.1089/ars.2010.3414

Dunbar, J., Krawczyk, K., Leem, J., Marks, C., Nowak, J., Regep, C., Georges, G., Kelm, S., Popovic, B., Deane, C.M., 2016. SAbPred: a structure-based antibody prediction server. Nucleic Acids Res. 44, W474-478. doi:10.1093/nar/gkw361

Eble N., J., Sauter, G., Epstein, J., Sesterhenn A., I., 2004. Tumors of the urinary system. In: World Health Organization classifi cation of tumours: pathology and genetics of tumours of the urinary system and male genital organs. Lyon Fr. IARC Press 89–123.

Eich, M.-L., Dyrskjøt, L., Netto, G.J., 2017. Toward personalized management in bladder cancer: the promise of novel molecular taxonomy. Virchows Arch. Int. J. Pathol. doi:10.1007/s00428-017-2119-x

Eijkelenboom, A., Burgering, B.M.T., 2013. FOXOs: signalling integrators for homeostasis maintenance. Nat. Rev. Mol. Cell Biol. 14, 83–97. doi:10.1038/nrm3507

Essers, M.A.G., Weijzen, S., de Vries-Smits, A.M.M., Saarloos, I., de Ruiter, N.D., Bos, J.L., Burgering, B.M.T., 2004. FOXO transcription factor activation by oxidative stress mediated by the small GTPase Ral and JNK. EMBO J. 23, 4802–4812. doi:10.1038/sj.emboj.7600476

Fan, W., Morinaga, H., Kim, J.J., Bae, E., Spann, N.J., Heinz, S., Glass, C.K., Olefsky, J.M., 2010. FoxO1 regulates Tlr4 inflammatory pathway signalling in macrophages. EMBO J. 29, 4223–4236. doi:10.1038/emboj.2010.268

Fernandez-Gomez, J., Madero, R., Solsona, E., Unda, M., Martinez-Piñeiro, L., Gonzalez, M., Portillo, J., Ojea, A., Pertusa, C., Rodriguez-Molina, J., Camacho, J.E., Rabadan, M., Astobieta, A., Montesinos, M., Isorna, S., Muntañola, P., Gimeno, A., Blas, M., Martinez-Piñeiro, J.A., 2009. Predicting nonmuscle invasive bladder cancer recurrence and progression in patients treated with bacillus Calmette-Guerin: the CUETO scoring model. J. Urol. 182, 2195–2203. doi:10.1016/j.juro.2009.07.016

Fitzmaurice, C., Barber, R.M., Barregard, L., Bhutta, Z.A., Brenner, H., Dicker, D.J., Chimed-Orchir, O., Dandona, R., Dandona, L., Fleming, T., Forouzanfar, M.H., Hancock, J., Hay, R.J., Hunter-Merrill, R., Huynh, C., Hosgood, H.D., Johnson, C.O., Jonas, J.B., Khubchandani, J., Kumar, G.A., Kutz, M., Lan, Q., Larson, H.J., Liang, X., Lim, S.S., Lopez, A.D., MacIntyre, M.F., Marczak, L., Marquez, N., Mokdad, A.H., Pinho, C., Pourmalek, F., Salomon, J.A., Sanabria, J.R., Sandar, L., Sartorius, B., Schwartz, S.M., Shackelford, K.A., Shibuya, K., Stanaway, J., Steiner, C., Sun, J., Takahashi, K., Vollset, S.E., Vos, T., Wagner, J.A., Wang, H., Westerman, R., Zeeb, H., Zoeckler, L., Abd-Allah, F., Ahmed, M.B., Alabed, S., Alam, N.K., Aldhahri, S.F., Alem, G., Alemayohu, M.A., Ali, R., Al-Raddadi, R., Amare, A., Amoako, Y., Artaman, A., Asayesh, H., Atnafu, N., Awasthi, A., Saleem, H.B., Barac, A., Bedi, N., Bensenor, I., Berhane, A., Bernabé, E., Betsu, B., Binagwaho, A., Boneya, D., Campos-Nonato, I., Castañeda-Orjuela, C., Catalá-López, F., Chiang, P., Chibueze, C., Chitheer, A., Choi, J.-Y., Cowie, B., Damtew, S., das Neves, J., Dey, S., Dharmaratne, S., Dhillon, P., Ding, E., Driscoll, T., Ekwueme, D., Endries, A.Y., Farvid, M., Farzadfar, F., Fernandes, J., Fischer, F., G/Hiwot, T.T., Gebru, A., Gopalani, S., Hailu, A., Horino, M., Horita, N., Husseini, A., Huybrechts, I., Inoue, M., Islami, F., Jakovljevic, M., James, S., Javanbakht, M., Jee, S.H., Kasaeian, A., Kedir, M.S., Khader, Y.S., Khang, Y.-H., Kim, D., Leigh, J., Linn, S., Lunevicius, R., El Razek, H.M.A., Malekzadeh, R., Malta, D.C., Marcenes, W., Markos, D., Melaku, Y.A., Meles, K.G., Mendoza, W., Mengiste, D.T., Meretoja, T.J., Miller, T.R., Mohammad, K.A., Mohammadi, A., Mohammed, S., Moradi-Lakeh, M., Nagel, G., Nand, D., Le Nguyen, Q., Nolte, S., Ogbo, F.A., Oladimeji, K.E., Oren, E., Pa, M., Park, E.-K., Pereira, D.M., Plass, D., Qorbani, M., Radfar, A., Rafay, A., Rahman, M., Rana, S.M., Søreide, K., Satpathy, M., Sawhney, M., Sepanlou, S.G., Shaikh, M.A., She, J., Shiue, I., Shore, H.R., Shrime, M.G., So, S., Soneji, S., Stathopoulou, V., Stroumpoulis, K., Sufiyan, M.B., Sykes, B.L., Tabarés-Seisdedos, R., Tadese, F., Tedla, B.A., Tessema, G.A., Thakur, J.S., Tran, B.X., Ukwaja, K.N., Uzochukwu, B.S.C., Vlassov, V.V., Weiderpass, E., Wubshet Terefe, M., Yebyo, H.G., Yimam, H.H., Yonemoto, N., Younis, M.Z., Yu, C., Zaidi, Z., Zaki, M.E.S., Zenebe, Z.M., Murray, C.J.L., Naghavi, M., 2017. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study. JAMA Oncol. 3, 524–548. doi:10.1001/jamaoncol.2016.5688

Forkhead boxes (FOX) Gene Family | HUGO Gene Nomenclature Committee [WWW Document], n.d. URL http://www.genenames.org/cgi-bin/genefamilies/set/508 (accessed 7.22.17).

Fridman, W.H., Pagès, F., Sautès-Fridman, C., Galon, J., 2012. The immune contexture in human tumours: impact on clinical outcome. Nat. Rev. Cancer 12, 298–306. doi:10.1038/nrc3245

Fu, W., Hu, W., Shi, L., Mundra, J.J., Xiao, G., Dustin, M.L., Liu, C.-J., 2017. Foxo4- and Stat3-dependent IL-10 production by progranulin in regulatory T cells restrains inflammatory arthritis. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. 31, 1354–1367. doi:10.1096/fj.201601134R

Furuyama, T., Nakazawa, T., Nakano, I., Mori, N., 2000. Identification of the differential distribution patterns of mRNAs and consensus binding sequences for mouse DAF-16 homologues. Biochem. J. 349, 629–634.

Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., Tosolini, M., Camus, M., Berger, A., Wind, P., Zinzindohoué, F., Bruneval, P., Cugnenc, P.-H., Trajanoski, Z., Fridman, W.-H., Pagès, F., 2006. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science 313, 1960–1964. doi:10.1126/science.1129139

Gee, J.R., Jarrard, D.F., Bruskewitz, R.C., Moon, T.D., Hedican, S.P., Leverson, G.E., Nakada, S.Y., Messing, E.M., 2009. Reduced bladder cancer recurrence rate with cardioprotective aspirin after intravesical bacille Calmette-Guérin. BJU Int. 103, 736–739. doi:10.1111/j.1464-410X.2008.08123.x

Gentles, A.J., Newman, A.M., Liu, C.L., Bratman, S.V., Feng, W., Kim, D., Nair, V.S., Xu, Y., Khuong, A., Hoang, C.D., Diehn, M., West, R.B., Plevritis, S.K., Alizadeh, A.A., 2015. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat. Med. 21, 938–945. doi:10.1038/nm.3909

GTEx Consortium, 2013. The Genotype-Tissue Expression (GTEx) project. Nat. Genet. 45, 580–585. doi:10.1038/ng.2653

Guo, S., Rena, G., Cichy, S., He, X., Cohen, P., Unterman, T., 1999. Phosphorylation of serine 256 by protein kinase B disrupts transactivation by FKHR and mediates effects of insulin on insulin-like growth factor-binding protein-1 promoter activity through a conserved insulin response sequence. J. Biol. Chem. 274, 17184–17192.

Hanahan, D., Coussens, L.M., 2012. Accessories to the crime: functions of cells recruited to the tumor microenvironment. Cancer Cell 21, 309–322. doi:10.1016/j.ccr.2012.02.022

Hanahan, D., Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. Cell 144, 646–674. doi:10.1016/j.cell.2011.02.013

Hedegaard, J., Lamy, P., Nordentoft, I., Algaba, F., Høyer, S., Ulhøi, B.P., Vang, S., Reinert, T., Hermann, G.G., Mogensen, K., Thomsen, M.B.H., Nielsen, M.M., Marquez, M., Segersten, U., Aine, M., Höglund, M., Birkenkamp-Demtröder, K., Fristrup, N., Borre, M., Hartmann, A., Stöhr, R., Wach, S., Keck, B., Seitz, A.K., Nawroth, R., Maurer, T., Tulic, C., Simic, T., Junker, K., Horstmann, M., Harving, N., Petersen, A.C., Calle, M.L., Steyerberg, E.W., Beukers, W., van Kessel, K.E.M., Jensen, J.B., Pedersen, J.S., Malmström, P.-U., Malats, N., Real, F.X., Zwarthoff, E.C., Ørntoft, T.F., Dyrskjøt, L., 2016. Comprehensive Transcriptional Analysis of Early-Stage Urothelial Carcinoma. Cancer Cell 30, 27–42. doi:10.1016/j.ccell.2016.05.004

Hedrick, S.M., Hess Michelini, R., Doedens, A.L., Goldrath, A.W., Stone, E.L., 2012. FOXO transcription factors throughout T cell biology. Nat. Rev. Immunol. 12, 649–661. doi:10.1038/nri3278

Hong, Y.M., Loughlin, K.R., 2008. Economic impact of tumor markers in bladder cancer surveillance. Urology 71, 131–135. doi:10.1016/j.urology.2007.08.014

Hori, S., Nomura, T., Sakaguchi, S., 2003. Control of regulatory T cell development by the transcription factor Foxp3. Science 299, 1057–1061. doi:10.1126/science.1079490

Huo, X., Liu, S., Shao, T., Hua, H., Kong, Q., Wang, J., Luo, T., Jiang, Y., 2014. GSK3 Protein Positively Regulates Type I Insulin-like Growth Factor Receptor through Forkhead Transcription Factors FOXO1/3/4. J. Biol. Chem. 289, 24759–24770. doi:10.1074/jbc.M114.580738

Jacobs, F.M.J., Heide, L.P. van der, Wijchers, P.J.E.C., Burbach, J.P.H., Hoekman, M.F.M., Smidt, M.P., 2003. FoxO6, a Novel Member of the FoxO Class of

Transcription Factors with Distinct Shuttling Dynamics. J. Biol. Chem. 278, 35959–35967. doi:10.1074/jbc.M302804200
Kaestner, K.H., Knöchel, W., Martínez, D.E., 2000. Unified nomenclature for the winged helix/forkhead transcription factors. Genes Dev. 14, 142–146. doi:10.1101/gad.14.2.142
Karanicolas, J., Corn, J.E., Chen, I., Joachimiak, L.A., Dym, O., Peck, S.H., Albeck, S., Unger, T., Hu, W., Liu, G., Delbecq, S., Montelione, G.T., Spiegel, C.P., Liu, D.R., Baker, D., 2011. A de novo protein binding pair by computational design and directed evolution. Mol. Cell 42, 250–260. doi:10.1016/j.molcel.2011.03.010
Karlsson, O.A., Ramirez, J., Oberg, D., Malmqvist, T., Engstrom, A., Friberg, M., Chi, C.N., Widersten, M., Trave, G., Nilsson, M.T.I., Jemth, P., 2015. Design of a PDZbody, a bivalent binder of the E6 protein from human papillomavirus. Sci. Rep. 5, 9382. doi:10.1038/srep09382
Kerdiles, Y.M., Stone, E.L., Beisner, D.R., Beisner, D.L., McGargill, M.A., Ch'en, I.L., Stockmann, C., Katayama, C.D., Hedrick, S.M., 2010. Foxo transcription factors control regulatory T cell development and function. Immunity 33, 890–904. doi:10.1016/j.immuni.2010.12.002
Kim, H.-J., Cantor, H., 2014. CD4 T-cell subsets and tumor immunity: the helpful and the not-so-helpful. Cancer Immunol. Res. 2, 91–98. doi:10.1158/2326-6066.CIR-13-0216
Kim, S.H., Miller, F.R., Tait, L., Zheng, J., Novak, R.F., 2009. Proteomic and phosphoproteomic alterations in benign, premalignant and tumor human breast epithelial cells and xenograft lesions: biomarkers of progression. Int. J. Cancer J. Int. Cancer 124, 2813–2828. doi:10.1002/ijc.24278
Kim, S.Y., Ko, Y.S., Park, J., Choi, Y., Park, J.-W., Kim, Y., Pyo, J.-S., Yoo, Y.B., Lee, J.-S., Lee, B.L., 2016. Forkhead Transcription Factor FOXO1 Inhibits Angiogenesis in Gastric Cancer in Relation to SIRT1. Cancer Res. Treat. Off. J. Korean Cancer Assoc. 48, 345–354. doi:10.4143/crt.2014.247
Kim, T.-H., Jo, S.-W., Lee, Y.S., Kim, Y.-J., Lee, S.-C., Kim, W.-J., Yun, S.J., 2009. Forkhead box O-class 1 and forkhead box G1 as prognostic markers for bladder cancer. J. Korean Med. Sci. 24, 468–473. doi:10.3346/jkms.2009.24.3.468
Klotz, L.-O., Sánchez-Ramos, C., Prieto-Arroyo, I., Urbánek, P., Steinbrenner, H., Monsalve, M., 2015. Redox regulation of FoxO transcription factors. Redox Biol. 6, 51–72. doi:10.1016/j.redox.2015.06.019
Kluth, L.A., Black, P.C., Bochner, B.H., Catto, J., Lerner, S.P., Stenzl, A., Sylvester, R., Vickers, A.J., Xylinas, E., Shariat, S.F., 2015. Prognostic and Prediction Tools in Bladder Cancer: A Comprehensive Review of the Literature. Eur. Urol. 68, 238–253. doi:10.1016/j.eururo.2015.01.032
Knowles, M.A., Hurst, C.D., 2015. Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. Nat. Rev. Cancer 15, 25–41. doi:10.1038/nrc3817
Kwon, I.-K., Wang, R., Thangaraju, M., Shuang, H., Liu, K., Dashwood, R., Dulin, N., Ganapathy, V., Browning, D.D., 2010. PKG inhibits TCF signaling in colon cancer cells by blocking beta-catenin expression and activating FOXO4. Oncogene 29, 3423–3434. doi:10.1038/onc.2010.91
Lam, E.W.-F., Brosens, J.J., Gomes, A.R., Koo, C.-Y., 2013. Forkhead box proteins: tuning forks for transcriptional harmony. Nat. Rev. Cancer 13, 482–495. doi:10.1038/nrc3539
Lapidoth, G.D., Baran, D., Pszolla, G.M., Norn, C., Alon, A., Tyka, M.D., Fleishman, S.J., 2015. AbDesign: An algorithm for combinatorial backbone design guided by natural conformations and sequences. Proteins 83, 1385–1406. doi:10.1002/prot.24779

Lee, M.-J., Yu, G.-R., Yoo, H.-J., Kim, J.-H., Yoon, B.-I., Choi, Y.-K., Kim, D.-G., 2009. ANXA8 down-regulation by EGF-FOXO4 signaling is involved in cell scattering and tumor metastasis of cholangiocarcinoma. Gastroenterology 137, 1138–1150, 1150.e1–9. doi:10.1053/j.gastro.2009.04.015

Lei, H., Quelle, F.W., 2009. FOXO Transcription Factors Enforce Cell Cycle Checkpoints and Promote Survival of Hematopoietic Cells after DNA Damage. Mol. Cancer Res. 7, 1294–1303. doi:10.1158/1541-7786.MCR-08-0531

Link, W., Fernandez-Marcos, P.J., 2017. FOXO transcription factors at the interface of metabolism and cancer. Int. J. Cancer. doi:10.1002/ijc.30840

Lloreta, J., Font-Tello, A., Juanpere, N., Frances, A., Lorenzo, M., Nonell, L., de Muga, S., Vázquez, I., Cecchini, L., Hernández-Llodrà, S., 2017. FOXO1 down-regulation is associated with worse outcome in bladder cancer and adds significant prognostic information to p53 overexpression. Hum. Pathol. 62, 222–231. doi:10.1016/j.humpath.2016.12.022

Löfblom, J., 2011. Bacterial display in combinatorial protein engineering. Biotechnol. J. 6, 1115–1129. doi:10.1002/biot.201100129

Luck, K., Charbonnier, S., Trave, G., 2012. The emerging contribution of sequence context to the specificity of protein interactions mediated by PDZ domains. Febs Lett. 586, 2648–2661. doi:10.1016/j.febslet.2012.03.056

Luo, C.T., Liao, W., Dadi, S., Toure, A., Li, M.O., 2016. Graded Foxo1 activity in Treg cells differentiates tumour immunity from spontaneous autoimmunity. Nature 529, 532–536. doi:10.1038/nature16486

Manning, B.D., Toker, A., 2017. AKT/PKB Signaling: Navigating the Network. Cell 169, 381–405. doi:10.1016/j.cell.2017.04.001

Maschberger, M., Breitsprecher, D., 2015. Rapid Quantification of Unfolded Proteins for Quality Control and Optimization of Storage Conditions [WWW Document]. NanoTemper Technol. URL https://nanotempertech.com/resources/rapid-quantification-unfolded-proteins-quality-control-optimization-storage-conditions/ (accessed 7.19.17).

Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim. Biophys. Acta 405, 442–451.

Medzhitov, R., 2008. Origin and physiological roles of inflammation. Nature 454, 428–435. doi:10.1038/nature07201

Misetic, V., 2016. Application Note NT-PR-007 - Unfolding without Tryptophan [WWW Document]. URL http://www.nanotemper-technologies.com/media/application-notes/nanodsf/ (accessed 9.26.16).

Mostofi, F.K., Sobin, L.H., Torloni, H., Organization, W.H., 1973. Histological typing of urinary bladder tumours. Who Chron. 28, 130-131

Myatt, S.S., Lam, E.W.-F., 2007. The emerging roles of forkhead box (Fox) proteins in cancer. Nat. Rev. Cancer 7, 847–859. doi:10.1038/nrc2223

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., Alizadeh, A.A., 2015. Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods 12, 453–457. doi:10.1038/nmeth.3337

Nordentoft, I., Lamy, P., Birkenkamp-Demtröder, K., Shumansky, K., Vang, S., Hornshøj, H., Juul, M., Villesen, P., Hedegaard, J., Roth, A., Thorsen, K., Høyer, S., Borre, M., Reinert, T., Fristrup, N., Dyrskjøt, L., Shah, S., Pedersen, J.S., Ørntoft, T.F., 2014. Mutational context and diverse clonal development in early and late bladder cancer. Cell Rep. 7, 1649–1663. doi:10.1016/j.celrep.2014.04.038

Obsil, T., Obsilova, V., 2011. Structural basis for DNA recognition by FOXO proteins. Biochim. Biophys. Acta 1813, 1946–1953. doi:10.1016/j.bbamcr.2010.11.025

Owens, B., 2017. Faster, deeper, smaller-the rise of antibody-like scaffolds. Nat. Biotechnol. 35, 602–603. doi:10.1038/nbt0717-602

Pan, C.-W., Jin, X., Zhao, Y., Pan, Y., Yang, J., Karnes, R.J., Zhang, J., Wang, L., Huang, H., 2017. AKT-phosphorylated FOXO1 suppresses ERK activation and chemoresistance by disrupting IQGAP1-MAPK interaction. Embo J. 36, 995–1010. doi:10.15252/embj.201695534

Pearce, L.R., Komander, D., Alessi, D.R., 2010. The nuts and bolts of AGC protein kinases. Nat. Rev. Mol. Cell Biol. 11, 9–22. doi:10.1038/nrm2822

Schneider, S., Buchert, M., Georgiev, O., Catimel, B., Halford, M., Stacker, S.A., Baechi, T., Moelling, K., Hovens, C.M., 1999. Mutagenesis and selection of PDZ domains that bind new protein targets. Nat. Biotechnol. 17, 170–175.

Shiota, M., Song, Y., Yokomizo, A., Kiyoshima, K., Tada, Y., Uchino, H., Uchiumi, T., Inokuchi, J., Oda, Y., Kuroiwa, K., Tatsugami, K., Naito, S., 2010. Foxo3a suppression of urothelial cancer invasiveness through Twist1, Y-box-binding protein 1, and E-cadherin regulation. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 16, 5654–5663. doi:10.1158/1078-0432.CCR-10-0376

Skrlec, K., Strukelj, B., Berlec, A., 2015. Non-immunoglobulin scaffolds: a focus on their targets. Trends Biotechnol. 33, 408–418. doi:10.1016/j.tibtech.2015.03.012

Su, B., Gao, L., Baranowski, C., Gillard, B., Wang, J., Ransom, R., Ko, H.-K., Gelman, I.H., 2014. A Genome-Wide RNAi Screen Identifies FOXO4 as a Metastasis-Suppressor through Counteracting PI3K/AKT Signal Pathway in Prostate Cancer. Plos One 9, e101411. doi:10.1371/journal.pone.0101411

Su, L., Liu, X., Chai, N., Lv, L., Wang, R., Li, X., Nie, Y., Shi, Y., Fan, D., 2014. The transcription factor FOXO4 is down-regulated and inhibits tumor proliferation and metastasis in gastric cancer. BMC Cancer 14, 378. doi:10.1186/1471-2407-14-378

Sykes, S.M., Lane, S.W., Bullinger, L., Kalaitzidis, D., Yusuf, R., Saez, B., Ferraro, F., Mercier, F., Singh, H., Brumme, K.M., Acharya, S.S., Scholl, C., Tothova, Z., Attar, E.C., Fröhling, S., DePinho, R.A., Gilliland, D.G., Armstrong, S.A., Scadden, D.T., 2011. AKT/FOXO Signaling Enforces Reversible Differentiation Blockade in Myeloid Leukemias. Cell 146, 697–708. doi:10.1016/j.cell.2011.07.032

Sylvester, R.J., van der Meijden, A.P.M., Oosterlinck, W., Witjes, J.A., Bouffioux, C., Denis, L., Newling, D.W.W., Kurth, K., 2006. Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. Eur. Urol. 49, 466-465; discussion 475-477. doi:10.1016/j.eururo.2005.12.031

Tian, L., Goldstein, A., Wang, H., Ching Lo, H., Sun Kim, I., Welte, T., Sheng, K., Dobrolecki, L.E., Zhang, X., Putluri, N., Phung, T.L., Mani, S.A., Stossi, F., Sreekumar, A., Mancini, M.A., Decker, W.K., Zong, C., Lewis, M.T., Zhang, X.H.-F., 2017. Mutual regulation of tumour vessel normalization and immunostimulatory reprogramming. Nature 544, 250–254. doi:10.1038/nature21724

Tothova, Z., Kollipara, R., Huntly, B.J., Lee, B.H., Castrillon, D.H., Cullen, D.E., McDowell, E.P., Lazo-Kallanian, S., Williams, I.R., Sears, C., Armstrong, S.A., Passegué, E., DePinho, R.A., Gilliland, D.G., 2007. FoxOs are critical mediators of hematopoietic stem cell resistance to physiologic oxidative stress. Cell 128, 325–339. doi:10.1016/j.cell.2007.01.003

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., Pachter, L., 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7, 562–578. doi:10.1038/nprot.2012.016

Tzivion, G., Dobson, M., Ramakrishnan, G., 2011. FoxO transcription factors; Regulation by AKT and 14-3-3 proteins. Biochim. Biophys. Acta 1813, 1938–1945. doi:10.1016/j.bbamcr.2011.06.002

Urbánek, P., Klotz, L.-O., 2016. Posttranscriptional regulation of FOXO expression: microRNAs and beyond. Br. J. Pharmacol. 174. doi:10.1111/bph.13471

Wakita, D., Sumida, K., Iwakura, Y., Nishikawa, H., Ohkuri, T., Chamoto, K., Kitamura, H., Nishimura, T., 2010. Tumor-infiltrating IL-17-producing gammadelta T cells support the progression of tumor by promoting angiogenesis. Eur. J. Immunol. 40, 1927–1937. doi:10.1002/eji.200940157

Wang, J.-H., Tang, H.-S., Li, X.-S., Zhang, X.-L., Yang, X.-Z., Zeng, L.-S., Ruan, Q., Huang, Y.-H., Liu, G.-J., Wang, J., Cui, S.-Z., 2017. Elevated FOXO6 expression correlates with progression and prognosis in gastric cancer. Oncotarget 8, 31682–31691. doi:10.18632/oncotarget.15920

Wang, S., Xia, P., Huang, G., Zhu, P., Liu, J., Ye, B., Du, Y., Fan, Z., 2016. FoxO1-mediated autophagy is required for NK cell development and innate immunity. Nat. Commun. 7. doi:10.1038/ncomms11023

Yamashita, H., Amponsa, V.O., Warrick, J.I., Zheng, Z., Clark, P.E., Raman, J.D., Wu, X.-R., Mendelsohn, C., DeGraff, D.J., 2017. On a FOX hunt: functions of FOX transcriptional regulators in bladder cancer. Nat. Rev. Urol. 14, 98–106. doi:10.1038/nrurol.2016.239

Zhang, W., Hietakangas, V., Wee, S., Lim, S.C., Gunaratne, J., Cohen, S.M., 2013. ER stress potentiates insulin resistance through PERK-mediated FOXO phosphorylation. Genes Dev. 27, 441–449. doi:10.1101/gad.201731.112